

University of Dundee

DOCTOR OF MEDICINE

A Study of the Consultation and Relational Empathy (CARE) Measure for Medical Students

Alnoman, Nashwan Nemat Hanna

Award date:
2014

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

DOCTOR OF MEDICINE

A Study of the Consultation and Relational Empathy (CARE) Measure for Medical Students

Nashwan Nemat Hanna Alnoman

2014

University of Dundee

Conditions for Use and Duplication

Copyright of this work belongs to the author unless otherwise identified in the body of the thesis. It is permitted to use and duplicate this work only for personal and non-commercial research, study or criticism/review. You must obtain prior written consent from the author for any other use. Any quotation from this thesis must be acknowledged using the normal academic conventions. It is not permitted to supply the whole or part of this thesis to any other person or to post the same on any website or other online location without the prior written consent of the author. Contact the Discovery team (discovery@dundee.ac.uk) with any queries about the use or acknowledgement of this work.

**A Study of the Consultation and Relational Empathy
(CARE) Measure for Medical Students**

Nashwan Nemat Hanna Alnoman

Thesis Submitted for the Degree of PhD

UNIVERSITY OF DUNDEE

May 2014

Dedication

To my dearest wife and lovely children

To my dad soul, my mother, brother and sister

To all who support & encourage

List of Content

Content	Page NO.
Dedication	i
List of Contents	ii
List of Figures	ix
List of Tables	x
List of Appendices	xii
Acknowledgements	xiii
Declaration	xiv
Summary of Thesis	xv
List of Abbreviations	xviii

1 Introduction	1
1.1 The importance of healthcare quality	1
1.1.1 ‘Better Health, Better Care’	1
1.1.2 ‘The Healthcare Quality Strategy for NHSScotland’	2
1.1.3 The person-centred approach	3
1.2 Rational of the study	6
1.3 Aim of thesis	7
1.4 Overview of the thesis	7
2 Literature review.....	10
2.1 Scope of the literature review	10
2.2 Sources used in this study	10
2.3 Development of the clinical consultation	13
2.4 The clinical consultation	14
2.4.1 Common models of the consultation	14

2.4.2	<i>Consultation skills and the doctor–patient relationship</i>	18
2.4.3	<i>The development of the doctor–patient relationship</i>	18
2.4.4	<i>The effect of the doctor–patient relationship</i>	20
2.4.5	<i>The doctor–patient relationship in medical education</i>	22
2.5	Empathy in the healthcare context	26
2.5.1	<i>Definition and historical background</i>	26
2.5.2	<i>Empathy and quality of healthcare</i>	29
2.5.3	<i>The effect of empathy on patient enablement and health outcome</i>	31
2.5.4	<i>Empathy in medical education</i>	34
2.5.5	<i>Empathy decline</i>	35
2.5.6	<i>The need for empathy measure</i>	38
2.6	Assessment in medical education	41
2.6.1	<i>General principles of assessment</i>	41
2.6.2	<i>Criteria for good assessment processes</i>	43
2.6.3	<i>Assessment in undergraduate medical education</i>	44
2.6.3.1	Miller’s model of assessment	44
2.6.3.2	The Cambridge Model	46
2.6.4	<i>Formative and summative assessments</i>	47
2.6.4.1	Combining formative and summative assessments	49
2.6.5	<i>The importance of feedback on consultation skills</i>	51
2.6.6	<i>Methods of feedback in medical education</i>	52
2.6.7	<i>Assessment as programme</i>	55
2.6.8	<i>Competency and outcome-based assessment</i>	56
2.6.8.1	Blueprinting	58
2.7	Key concepts in assessment	59
2.7.1	<i>Utility of assessment</i>	59
2.7.2	<i>The validity of the assessment tools</i>	60
2.7.3	<i>The reliability of assessment tools</i>	63

2.7.4	<i>Types of reliability</i>	65
2.7.5	<i>Reducing measurement error and improving reliability.....</i>	69
2.7.6	<i>The relationship between reliability and validity of assessment tools</i>	71
3	Systematic review of empathy measures for health professionals .	73
3.1	Introduction	73
3.2	Background and rationale	73
3.3	The aim of the systematic review in this thesis	75
3.4	Search strategy	76
3.4.1	<i>Selection criteria</i>	76
3.4.2	<i>Search procedure</i>	77
3.4.3	<i>Data extraction.....</i>	80
3.4.4	<i>Synthesis.....</i>	84
3.4.5	<i>Limitations</i>	84
3.5	Measures of empathy for health professionals	86
3.5.1	<i>Self-report empathy measures.....</i>	86
3.5.1.1	Jefferson Scale of Physician Empathy	86
3.5.1.2	Hogan's Empathy Scale	88
3.5.1.3	The Interpersonal Reactivity Index	90
3.5.1.4	The Empathy Communication Skills Scale	92
3.5.1.5	The Balanced Emotional Empathy Scale.....	93
3.5.2	<i>Patient-rating empathy measures</i>	94
3.5.2.1	The Consultation and Relational Empathy Measure	94
3.5.3	<i>Observer-rating empathy measures</i>	96
3.5.3.1	The Four Habits Coding Scheme	96
3.5.3.2	Empathic Communication Coding System.....	98
3.5.4	<i>Combined empathy measures.....</i>	99
3.5.4.1	The Reynolds Empathy Scale	99
3.5.4.2	The Barrett-Lennard Relationship Inventory.....	100
3.5.4.3	Jefferson Scale of Patient's Perception of Physician Empathy.....	102

3.5.4.4	Empathy Construct Rating Scale	102
3.6	Results of the systematic review	104
3.7	Choosing a suitable measure for this study	107
4	Methods used for the utilisation of the CARE measure.....	109
4.1	The study design	109
4.2	Ethics Consideration and approval	110
4.2.1	<i>Patients' ethical issues</i>	111
4.2.2	<i>Students' ethical issues</i>	111
4.3	Background on the CARE measure.....	112
4.3.1	<i>Development of the CARE measure</i>	112
4.3.2	<i>Preliminary validation of the CARE measure</i>	113
4.3.3	<i>Current use of the CARE measure</i>	115
4.3.4	<i>International use of the CARE measure</i>	116
4.3.5	<i>The CARE questionnaire for medical students</i>	118
4.4	Outline of the Dundee Medical School programme	120
4.5	Participants characteristics and instruction forms	122
4.5.1	<i>Students</i>	122
4.5.2	<i>Patients</i>	122
4.5.3	<i>Teaching staff, tutors, and practice staff</i>	123
4.6	Data collection.....	125
4.6.1	<i>Student recruitment (introductory briefing)</i>	125
4.6.2	<i>Data management</i>	126
4.6.3	<i>Gathering students' views</i>	126
4.6.4	<i>Gathering and utilising patients' comments</i>	127
4.6.5	<i>Provision of feedback to students</i>	128
4.7	Statistical analysis.....	129
4.8	Summary	133

5 Results of the CARE Measure Utility	134
5.1 Introduction	134
5.2 Descriptive statistics	134
5.2.1 <i>Student sample and response rate.....</i>	<i>134</i>
5.2.2 <i>Reasons for non-participating</i>	<i>136</i>
5.2.3 <i>Reasons for not completing the minimum number of questionnaires ...</i>	<i>137</i>
5.2.4 <i>Missing data.....</i>	<i>138</i>
5.3 Descriptive analysis	140
5.3.1 <i>Mean and standard deviation of CARE measure scores.....</i>	<i>140</i>
5.3.2 <i>Distribution of students' means and confidence intervals</i>	<i>141</i>
5.3.3 <i>Students' mean scores distribution</i>	<i>142</i>
5.3.4 <i>A comparison of student and physician CARE score outcomes.....</i>	<i>142</i>
5.4 Qualitative results.....	144
5.4.1 <i>Patients' feedback (Comments)</i>	<i>144</i>
5.4.1.1 <i>Positive patient comments.....</i>	<i>145</i>
5.4.1.2 <i>Negative patient comments</i>	<i>147</i>
5.5 Quantitative results	149
5.5.1 <i>Reliability of the CARE measure with Students</i>	<i>149</i>
5.5.2 <i>Validity of the CARE measure with students.....</i>	<i>150</i>
5.6 Summary	152
6 Discussion of the CARE Measure Utility.....	153
6.1.1 <i>The method of CARE measure utilisation for medical students.....</i>	<i>153</i>
6.1.2 <i>Effect of Student response rate on the CARE measure utilisation.....</i>	<i>154</i>
6.1.3 <i>Why use 25 CARE questionnaire forms?</i>	<i>155</i>
6.1.4 <i>Reasons for not completing the minimum number of questionnaires ...</i>	<i>156</i>
6.1.5 <i>Reporting and handling missing data</i>	<i>158</i>
6.1.6 <i>Interpreting students' scores.....</i>	<i>160</i>

6.1.7	<i>Student mean scores</i>	161
6.1.8	<i>Student means versus GP means</i>	162
6.1.9	<i>Reliability of the CARE measure for students</i>	163
6.1.10	<i>Validity of the CARE measure for students</i>	165
6.1.11	<i>Reliability and educational impact</i>	166
6.1.12	<i>Patient, student, and staff feedback</i>	167
6.2	Strengths and limitations of this study	170
6.2.1	<i>Strengths</i>	170
6.2.2	<i>Limitations</i>	170
6.2.3	<i>Summary</i>	172
7	Mapping exercise	173
7.1	Background of mapping exercises	173
7.2	Introduction and rationale	173
7.3	Aims	176
7.4	Methods	176
7.4.1	<i>Study design and ethical issues</i>	176
7.4.2	<i>Questionnaire used in the mapping exercise</i>	176
7.4.3	<i>Piloting the mapping exercise</i>	179
7.4.4	<i>Exercise sample and data collection</i>	180
7.4.5	<i>Statistical analysis of the mapping exercise</i>	181
7.5	Results of the mapping exercise	182
7.5.1	<i>Response rate and level of agreement</i>	182
7.5.2	<i>Coverage of GMC attributes by assessment tools</i>	182
7.5.3	<i>The CARE measure in the Mapping Exercise</i>	184
7.6	Discussion and conclusion of the mapping exercise	188
7.6.1	<i>General review of Dundee Medical School assessment programme</i>	188
7.6.2	<i>The role of the CARE measure within the assessment programme</i>	189

8 Implications and Conclusions	191
8.1 Implications.....	191
8.1.1 <i>Implications for practice and medical education.....</i>	<i>191</i>
8.1.2 <i>Implications of patient feedback on student learning</i>	<i>193</i>
8.1.3 <i>Implications for the Dundee Medical School assessment programme .</i>	<i>195</i>
8.1.4 <i>Implications for future research</i>	<i>195</i>
8.2 Conclusions	197
 References	 199
 Appendices	 226

List of Figures

Figure Title	Page
Figure 2.1: The Calgary–Cambridge approach in consultations.	17
Figure 2.2: Miller’s pyramid.	45
Figure 2.3: The Cambridge Model.	47
Figure 2.4. The three-circle model for outcome-based education	57
Figure 3.1: Flow diagram for the search process.	79
Figure 5.1: The average scores of each question of CARE questionnaire versus the overall mean.	140
Figure 5.2: Students’ mean CARE measure scores distribution (histogram of means).	142
Figure 5.3: Difference between the student and GP mean CARE measure scores.	143

List of Tables

Table Title	Page
Table 2.1: New feedback delivery compared to the standard feedback methods.	54
Table 2.2: The types of reliability with the sources of differences (facets).	68
Table 3.1: Empathy assessment measures (Jefferson Scale of Physician Empathy (JSPE); Hogan's Empathy Scale (HES); Interpersonal Reactivity Index (IRI); Empathy Communication Skills Scale (ECSS); Balanced Emotional Empathy Scale (BEES); Consultation and Relational Empathy Measure (CARE); Reynolds Empathy Scale (RES); Barrett-Lennard Relationship Inventory (BLRI); Jefferson Scale of Patient's Perception of Physician Empathy (JSPPE); Empathy Construct Rating Scale (ECRS); Four Habits Coding Scheme (4HCS); Empathic Communication Coding System (ECCS)) and the studies included for each target population.	81
Table 4.1 Types of reliability and facets of the CARE measure for medical students.	131
Table 5.1: The distribution of the study sample according to the students' response rate.	135
Table 5.2: The distribution of participants according to completion of minimum required number of questionnaire forms.	136
Table 5.3: Numbers and percentages of non-valued data per each question of the CARE measure.	139
Table 5.4: The CARE measure means scores, standard deviation (SD) and 95% confidence intervals.	141
Table 5.5: The main categories of patients' comments.	145
Table 5.6: Specific and non-specific positive patients' comments	145

Table 5.7: The specific categories of the negative patients' comments.	147
Table 5.8: The CARE measure reliabilities according to number of patients.	150
Table 5.9: Correlation of CARE students' mean with other assessments of same construct.	151
Table 7.1: The CARE mapping exercise- Phase 3 assessment tools versus GMC attributes (n=17).	186

List of Appendices

- Appendix 1 Search procedure for the systematic review
- Appendix 2 Ethical approval letter
- Appendix 3 The original CARE questionnaire (doctors' version)
- Appendix 4 The CARE questionnaire (Students' version)
- Appendix 5 Instruction to patients
- Appendix 6 Feedback results to students (email), 2nd & 3rd years
- Appendix 7 Feedback results (means and confidence intervals)
- Appendix 8 Dundee GP block (mini-CEX) form
- Appendix 9 List of Positive patients' comments
- Appendix 10 List of Negative patients' comments
- Appendix 11 Email of CARE mapping exercise sent to participants
- Appendix 12 The CARE measure mapping exercise
- Appendix 13 Instructions on cover page of mapping exercise

Acknowledgements

I would like to express my sincere gratitude and thanks to the Iraqi Ministry of Higher Education and Scientific Researches for providing full funding for my scholarship. Also, I would like to thank the University of Dundee for the opportunity to study, do research, teach, and attend conferences and seminars.

I am grateful to my supervisors Dr Johnathan Dowell and Dr Douglas Murphy. Throughout this journey, you have provided me with wisdom and inspiration. Thank you so much for your never-ending patience and encouragement to get this work finished on time.

Special thanks are due to all my colleagues and staff at Dundee School of Medicine, Division of Clinical and Population Sciences and Education, who gave me support all the way. I would also like to acknowledge all participants involved in this research; in particular, the students and patients.

I am grateful to my friend Elizabeth Farrell for her kind help in proofreading.

Finally, great thanks to my family for their endurance and support through the difficult period of my life.

Thanks to all of you.

Nashwan Alnoman

May 2014

Declaration

I certify that the work presented in this thesis is, to the best of my knowledge and belief, original, except as acknowledged in the text. All references cited have been consulted and that the material has not been submitted, either in whole or in part, for a degree at this or any other university.

Name: Nashwan Nemat Hanna Alnoman

Signature:-----

Date:-----

I certify that Nashwan Nemat Hanna Alnoman has completed the equivalent of nine terms of experimental research and that he has fulfilled the conditions of the relevant Ordinance and Regulations of the University of Dundee, so that he is qualified to submit this thesis in application for the degree of Doctor of Philosophy.

Name: Johnathan Dowell

Signature:-----

Date:-----

Summary

Patients who are treated empathically tend to be more satisfied and have better health outcomes. Many instruments for measuring empathy in different health contexts have been developed for postgraduate and to a lesser extent undergraduate health professional use. This thesis presents the results of three studies. The first is a systematic review of the literature describing empathy measures for use in the health professional context. The review was conducted to find the best measure specifically designed for and evaluated in the clinical context that assessed empathy reliably and preferably from a patient perspective. The Consultation and Relational Empathy (CARE) measure was selected from the 12 potential empathy measures available. This instrument was developed and tested in general practice consultations in Scotland, and its usefulness for assessing physicians' empathy has been demonstrated.

The second study, which forms the main research component of the thesis, is an evaluation of the utility of the CARE measure in assessing medical undergraduates on general practice placements. This study aimed to establish whether the CARE measure offers a reliable, valid, and practical means of assessing certain aspects of senior clinical medical students' consultation skills. Fifth-year students who performed a two- or three-month GP attachment in academic years 2009–2010, 2010–2011, and 2011–2012 were invited to participate. In 2011 participation in the study became a course requirement. Consulted patients were invited to complete a modified CARE questionnaire 'on the student'. The form initially included 11 questions answered on a five-point scale; from the academic year 2010–2011, the form also included a free-text comment option. Patients completed questionnaires anonymously. Forms were collated

and forwarded to the principle investigator. A minimum of 25 forms were required for each student in order to minimise potential selection bias in the analysis.

In the first year of the study the response rate was 56 %. During the next two academic years, this rate improved to over 95%. The total number of the CARE measure forms returned was 2,145, with 1,808 adequately completed. The average score of the students recruited was 4.57 (standard deviation = 0.19), which is higher than average scores for general practitioners. Generalisability theory calculations showed that the CARE measure had high internal consistency, with 25 patients per student delivering an overall reliability coefficient of 0.74. Using the CARE measure with students significantly correlated at a moderate level with relevant outcome assessments such as the mini-Clinical Evaluation Exercise (mini-CEX), Objective Structured Clinical Examination (OSCE) at the end of the GP block, and tutor scores ($r = 0.38$, $r = 0.33$, $r = 0.33$, respectively; $P < 0.05$). The findings suggest that the CARE measure is a useful tool for assessing senior medical students' consultation skills, at least in the context of extended general practice attachments. In addition, the measure provides students with valuable and authentic patient feedback for their training.

The third study was a 'mapping exercise' designed to establish the role of current assessment tools at the University of Dundee in relation to *Tomorrow's Doctors* outcomes, and specifically where/how CARE might fit in. Dundee medical faculty staff, block leaders, undergraduate medical educators, and others who engaged directly with assessments and with experience of the existing exams were invited to 'map' the value of each aspect of the medical school's 11 assessment tools against General Medical Council (GMC) attributes. Data was collected electronically from 17 staff

members. The results of the exercise revealed both gaps in assessment and over-testing of specific GMC attributes. These findings offer Dundee Medical School an opportunity to optimise their approach to assessment and achieve a broader coverage of GMC outcomes. Thus, the CARE measure offers a logical extension to aspects of the OSCE assessment.

The results show that the CARE measure is a feasible, acceptable, reliable, and valid measure for the assessment of students. In addition, as it incorporates authentic patient feedback into training, this measure can offer unique utility to the medical school's curriculum and assessment as it incorporates the rarely employed patient voice.

List of Abbreviation

ANOVA	Analysis of variance
CARE	Consultation and Relational Empathy
EMI	Extended Matching Item
GMC	General Medical Council
GP	General Practitioner
JSPE	Jefferson Scale of Physician Empathy
KCS	Kalamazoo Consensus Statement
Mini-CEX	Mini-Clinical Evaluation Exercise
NHS	National Health Service
NQMC	National Quality Measure Clearinghouse
OSCE	Objective Structured Clinical Examination
PEI	Patient Enablement Instrument
PMETB	Postgraduate Medical Education and Training Board
RCGP	Royal College of General Practitioners

1 Introduction

This chapter describes the importance of quality in healthcare in general, with reference to the recent healthcare quality strategy for Scotland. It then focuses on the person-centred approach as one key requirement for high-quality healthcare. This chapter explains the rationale behind conducting the study and sets out the general aims. Finally, an outline of the thesis is presented.

1.1 The importance of healthcare quality

Provision of healthcare is not enough; a high level of efficiency and a high degree of service quality must also be achieved, given financial constraints and patient expectations in the modern context. This section will focus on the quality of healthcare in Scotland and describe the most recent strategies called for by NHSScotland in the last few years and the need for synchronous changes in medical education as a result.

1.1.1 ‘Better Health, Better Care’

This action plan was developed by the Scottish government after consultation with people who expressed their needs and expectations about the NHS. It was launched in December 2007 to improve health and the quality of health care services for all people in Scotland and achieve better partnership in the health care services. The aim of the ‘Better Health, Better Care’ action plan was to set out a proposal to create a mutual NHS, thought necessary to provide person-centred healthcare, for Scotland. In theory, people will then have a better chance to control their health and NHS services going forwards (NHSScotland, 2008).

‘Better Health, Better Care’ describes six dimensions of care which are essential for health care quality improvement: person centred, safe, effective, efficient, equitable, and timely.

1.1.2 ‘The Healthcare Quality Strategy for NHSScotland’

Since ‘Better Health, Better Care’ was published, many significant achievements such as improved waiting times, safety of patient in hospitals, and tackling health care associated infection have taken place (NHSScotland, 2010). Developments following that plan led to the publication of ‘The Healthcare Quality Strategy for NHSScotland’, which provides a new, important approach that puts the patient at the centre of everything, so that Scotland would become a world leader in healthcare quality (NHSScotland, 2010). The strategy described in this publication enables people to comment on their experiences of healthcare, share effectively in decision making regarding their care, and thus become more confident in the healthcare system. The belief underpinning this strategy is that the NHS will improve through healthcare providers listening to patient experiences and that patient will have an important role in monitoring their quality of care. In addition, the strategy gives impetus for the NHS to prepare personalised care plans both for people with special needs and for complex cases.

The healthcare requirements of most people in Scotland can be summarised as follows: to have caring and compassionate staff, clear communication, effective collaboration, a clean and safe care environment, continuity of care, and a high-quality clinical experience. Focusing on these six dimensions of quality will improve and raise the

quality of healthcare to a world-leading standard, and thus the aim of the strategy is to achieve the following three key ‘Quality Ambitions’ (NHSScotland, 2010):

- 1) Person-centred healthcare (enabling mutually beneficial partnerships between patients, their families, and healthcare providers)
- 2) Safe healthcare (ensuring no avoidable injury or harm from healthcare occurs and providing an appropriate, clean, and safe healthcare environment)
- 3) Clinically effective healthcare (ensuring that appropriate healthcare services and support are provided at the right time to everyone who will benefit from them and eliminating wasteful and/or harmful variations)

Of course, people expect safe healthcare that is provided without harm or injury and delivered in safe and clean environment, and health services in general should be clinically effective. However, this thesis focuses on the first quality ambition: the provision of person-centred care, as a patient’s perceived experience of care is as closely related with the way in which care is provided as it is with the kind of care that is given. In addition, there is evidence that, beyond patient satisfaction, the outcome of care improves when a person- or patient-centred approach is used. Hence, development, measurement, and study of person-centred approaches are critical for the creation and maintenance of the type of health service sought in Scotland.

1.1.3 The person-centred approach

This widely adopted programme in the UK puts people (patients and their families) at the centre of care, building caring relationships between patients and staff, aiming not only for better outcomes and experiences for patients and carers but also better job

satisfaction for staff. Also, this approach provides patients with an opportunity to share in planning, developing, and assessing how their own care meets their needs (NHSScotland, 2010). Historically, the concept of person-centred care arose within the primary care arena but was primarily used with the elderly, particularly those with dementia. In general practice, the person-centred approach was first suggested by Ian McWhinney from the University of Western Ontario, Canada (McWhinney, 1985, McWhinney, 1996, McWhinney and Freeman, 2009) and then highlighted by the development of the National Care Standards for Scotland (Scottish Executive, 2002). The person-centred approach is essential for delivering high-quality healthcare (Stewart, 1995); (Coyle and Williams, 2001); (Mead and Bower, 2000) and there is an increased emphasis on this approach in the recent literature (Dawood, 2005); (McCormack, 2003).

The phrase ‘patient-centred’ appeared for the first time in GMC documents in the second edition of *Tomorrow’s Doctors* (GMC, 2002). However, it was clear even in the first edition (1993) that medical schools were expected to teach students how to communicate with patients and that this element of patient-centred care should be part of the curriculum. Nonetheless, there is more to patient-centred care than communicating clearly and being considerate, and the evidence base has evolved around a number of closely related approaches, perhaps most clearly described by Stewart (Stewart, 1995). In essence, they describe clinical encounters as patient-centred when they deal effectively with patients’ understanding of their illness, beliefs about cause and treatment, expectations of care, feelings, and fears.

There are six interconnecting components of the patient-centred method: first, exploring the disease, examining the illness experience, and understanding the whole person in his/her particular context; then finding common ground to incorporate

prevention, assist health promotion, and enhance the patient-doctor relationship in a realistic manner. Thus, patient-centred care reflects an overall approach that encompasses an ethos as well as a number of steps and skills. There is no a standard package or approach that can be delivered to all patients, and the practice of patient-centred care is highly contextual, depending on the illness, patient, and healthcare system.

However, assessing such an individual and experiential aspect of care provision is challenging. Patient opinion is essential, and monitoring and responding to patient opinion is increasing both within the service and more broadly. For instance, the involvement of patients in the assessment of undergraduate medical students is now encouraged by the GMC as part of this process. According to Standard 16 of the Postgraduate Medical Education and Training Board Standards for Curricula,

‘There will be lay and patient input in the development and implementation of assessments.’

(GMC, 2010a); (Southgate and Grant, 2004)

But what ‘input’ from patients will help determine if care has been patient-centred or if a practitioner has demonstrated the required approach and skills? It is argued in this thesis that the concept of empathy is a key and practical construct which offers promise for this purpose.

1.2 Rational of the study

Research has demonstrated that patients who are treated empathically tend to be more satisfied, adhere to treatment recommendations more closely, and have better health outcomes. Among organisations that oversee the training and certification of doctors, there is now widespread acceptance of empathy as a vital professional skill, but there is not yet a consensus on how best to measure this skill. Therefore, the literature describing empathy in the doctor–patient relationship and the assessment of consultation skills was reviewed and summarised.

This revealed that many instruments to measure empathy in different health contexts have been developed for postgraduate and, to a lesser extent, undergraduate health professionals. In addition, some tools have been validated for assessing health professionals' empathy in practice. In this study, we examined whether measures designed for the clinical context could also be used to assess medical students' empathy during interactions with patients. Such an assessment would both improve medical students' training and enable faculty to evaluate students on an important aspect of being an effective doctor. A key premise underlying this study was the assumption that assessment of empathy is not currently being conducted to a satisfactory extent during medical training; therefore, to determine whether this assumption was accurate, we also carried out a mapping exercise of the tools currently available at the University of Dundee for assessing empathy, to assess what role such tools might play within the broader assessment framework.

1.3 Aim of thesis

The overall purpose of this thesis was to establish the utility of an instrument to measure medical students' ability to provide empathic care, in order to improve medical training. To achieve this aim, it was necessary to conduct three distinct studies:

1. a systematic review of the available tools for assessing empathy, followed by selection of the most promising tool for further testing in the context of UK undergraduate medical education;
2. an investigation to determine whether the selected tool would offer a reliable, valid, and practical means of assessing relevant skills and behaviours; and
3. an examination of the existing assessment framework to ascertain appropriate placement of the new tool within it.

1.4 Overview of the thesis

This thesis contains eight chapters. The first chapter is the introduction, which provides the context for this thesis. In this chapter, the importance of doctor–patient communication and empathy to healthcare quality is discussed, as well as the current drive towards increasingly patient-centred approaches.

Chapter 2 reviews common models of clinical consultation, the development of consultation skills as a subject area, key consultation skills with a focus on empathy in the patient–doctor relationship, and patient enablement and health outcomes. Empathy among medical students is explored, including the teaching and measuring of this clinical skill within the field of medical education. In addition, the ways in which

levels of empathy may decline in certain situations is discussed. The methods of assessment in medicine are presented in this chapter, and recent approaches to assessing undergraduate medical students' consultation skills are demonstrated, with a focus on the key concepts of validity and reliability in assessment.

Chapter 3 presents a systematic review of empathy measures, an outline of the search strategy with a list of selection criteria for publications, and the search procedure. Each empathy measure that has been tested in a health context is described, with details on each tool. This is followed by the results and a discussion of the process employed to select the CARE measure as the one considered most suitable for this study.

Chapter 4 provides a description of the methods used for testing reliability and validity of the CARE measure and outlines the research design, data collection, analytic approach, and ethical considerations. The reasons for the adoption of the CARE measure are expanded and changes made to the original CARE assessment form are described within a more detailed description of the literature concerning its development. The participants' characteristics and instructions to students, patients, and practices staff are described. This chapter also delineates data collection, handling, and preparation for statistical analysis.

Chapter 5 describes the findings of this study on the utility of the CARE questionnaire, particularly its reliability, acceptability, and feasibility. The chapter begins with a presentation of the descriptive statistics for study samples (mean, standard deviation, etc.) and plots thereof. This is followed by an analysis of the quantitative data to find the correlation and reliability of the CARE measure. The latter part of this chapter deals with the qualitative aspects of the study and expands on the acceptability, potential uses, and educational impact of the CARE measure.

Chapter 6 first considers the strengths and weaknesses of the evaluation of the CARE measure presented here. The chapter then discusses the findings, in the same sequence as the findings were presented in the Chapter 5, and provides an interpretation of what the findings might mean in the broader context. In addition, the methodological strengths and possible limitations of this study are highlighted and discussed.

Chapter 7 provides a detailed description of the mapping exercise conducted for the present study. The chapter begins by exploring the background of mapping exercises in research and the aim of the exercise and then goes on to describe the methods used to collect the data from participants and the plan for statistical analysis. Finally, this chapter concludes with a detailed description of the findings and an interpretation of those findings.

In Chapter 8, the broader implications of the thesis are discussed in terms of the potential effect of the use of the CARE measure on practice, medical education, and student learning, as well as on future research. In addition, the chapter summarises the final conclusions and key messages arising from this thesis.

2 Literature review

The chapter starts by describing the areas of literatures reviewed and outlines the searches used. It explores the process of the clinical consultation, the patient–doctor relationship, and the roles of the patient and doctor in the search for better healthcare quality. Relevant literature on the importance of empathy in the healthcare context is then explored, to assess the role of this specific construct within the consultation, the patient–doctor relationship, and quality of care. The literature on empathy in medical education is then outlined before the measurement of empathy in doctors is discussed. Finally, prior to focusing on empathy measurement in medical education, the chapter explores general aspects of assessment in medical education and reviews key concepts in assessment such as the reliability and validity of assessment tools.

2.1 Scope of the literature review

The literature review presents the following main areas:

1. the development of clinical consultation methods and the evolution of doctor–patient relationships;
2. the role of empathic consultation skills in person-centred approaches;
3. the need for measuring empathy in the context of healthcare;
4. the principles and methods of assessment in medical education;
5. models of assessment in medicine, and important key concepts for assessment.

2.2 Sources used in this study

The ‘CrossSearch’ service was used to search many different information sources, including databases, library catalogues, collections of electronic journals and e-books.

The following multidisciplinary indexes were searched for publications to analyse article bibliographies and track citations:

SCOPUS

Web of Knowledge

Zetoc

ATHENS

Google Scholar

In addition, other Internet resources were used, such as:

ScienceDirect

PubMed

University of Dundee Library Catalogue

Online searches of electronic books were also used, as well as some books or documents not yet available in electronic form. Furthermore, a Zetoc email alert was established in June 2009 to find newly published papers on the topic. Documents identified through these searches were examined, and relevant ones were included in this review. These search methods were supplemented with ‘reference chaining’, that is, the reference lists of retrieved papers were searched to identify any other relevant studies. Additionally, relevant articles were also identified by browsing the reference sections of new publications from the following related key journals:

Medical Education

Medical Teacher

Academic Medicine

British Medical Journal

Databases were searched using Boolean operators for combinations of the following key search terms: ‘consultation skills’, ‘clinical consultation’, ‘assessment tool reliability’, ‘patient satisfaction’, ‘medical students’, ‘patient-centred care’, ‘patient feedback’, ‘relational empathy’, and ‘empathy measure’. These terms were entered, along with truncated versions and equivalent terms in the Medical Subject Headings browser (also known as the MeSH browser) to ensure a broad sweep for relevant papers. Searches were limited to English-language articles.

Publications were searched from 1985 to 2013; except for a few older key books or articles, thereafter publications were identified via the ongoing alerts set up as described above. The literature search process was repeated yearly (most recently, September 2013) to identify new relevant articles. The bibliographic management tool ‘EndNote X4’ was used to store citations, papers, and annotations.

2.3 Development of the clinical consultation

The consultation or medical interview in clinical practice is the primary opportunity to explore patients' symptoms and illness and to seek related problems or concerns. It has been described as a 'meeting between experts', where the doctor is the expert in diagnosis and disease, and the patient is the expert in his/her own symptoms and experiences (Tuckett et al., 1985). The success of a consultation depends on its context. Priorities will differ between acute or emergency situations and straightforward diagnostic or therapeutic encounters, which include more psychosocial interactions. In the former context, efficient accurate history taking and information gathering that lead (ideally) to a precise diagnosis may be key (Silverman et al., 2005); (Epstein et al., 2008). In the latter situation, it may be vital to establish a highly supportive environment that encourages patients to raise their ideas and concerns. The consultation as a key part of the therapeutic process has been studied most in the primary care setting and there are a number alternative models or approaches that have been proposed.

2.4 The clinical consultation

For a long time, the consultation has been the vehicle for the art or craft of medical practice. After World War II, medicine and healthcare expanded in most western countries; medicine became more and more technically advanced, and large hospitals and super-specialised clinical care were part of this trend. However, due to experiences from clinical practice, a humanistic countermovement started among British general practitioners in the 1950s. Michael Balint's *The Doctor, his Patient, and the Illness* was one of the most significant contributions from this early period and introduced a psychodynamic perspective to GPs' encounters (Balint, 1957).

Medicine is more than the application of biomedical science. Practicing medicine also means trying to understand the patient's experiences as a unique person, in a psychosocial and existential context. Thus, consultation in clinical practice or the medical interview is an opportunity to explore patients' illness and reveal any related problems or concerns. It has been described as a 'meeting between experts', where the doctor is the expert in diagnosis and disease, and the patient is the expert in their own symptoms and experiences (Tuckett et al., 1985). The success of a consultation depends mostly on history taking and gathering information from patients, information that will lead, ideally, to a final precise diagnosis (Epstein et al., 2008). This topic has been most studied in the primary care setting and there are a number alternative models or approaches that have been proposed.

2.4.1 Common models of the consultation

During the Bayer–Fetzer Conference on Physician–Patient Communication in Medical Education in May 1999, different models of the consultation were examined and a list of the essential tasks of the consultation was produced, called the Kalamazoo

Consensus Statement (Bayer–Fetzer, 2001). Among the models presented were the following:

Helmann’s folk model: This model centres on the patient’s story and looks at the consultation from the patient’s perspective, using empathy to address the patient’s question (e.g. ‘I can see how sad/frustrated/angry/frightened/upset you are’) (Helmann, 1981). It was interesting because of its novel, at the time, focus on the patient’s perspective, beliefs, and concerns. However, it is not specific or detailed enough to work well as a clinical framework.

The Pendleton model: As the first structured modern consultation model, this seven task, patient-centred process consists of defining the patient’s true agenda, considering other problems, choosing an appropriate action for each problem, achieving a shared understanding, using time and resources appropriately, and establishing and maintaining a relationship with the patient. Examples of questions from this model include ‘Was there anything else you were hoping to discuss today?’, ‘What is your main fear/worry/concern about this problem?’, and ‘What were you hoping to get out of today?’ (Pendleton et al., 1984).

The Neighbour model (The inner consultation model): This model presents a more flexible but clearly related five-point system of connecting, summarising, handing over, ‘safety netting’, and housekeeping, alongside an awareness of ‘minimal cues’ (verbal and non-verbal) to help discover the unspoken agenda. Summarising is a useful tool to clarify the understanding of the problem and can also be used to demonstrate empathy (Neighbour, 1987).

The Calgary–Cambridge approach: This model, which now seems to be the most influential (Kurtz and Silverman, 1996), and defines the main stages of any consultation as 1) initiating the session, 2) gathering information, 3) physical examination, 4) explanation and planning, and 5) closing the session (Figure 2.1). Unlike earlier models, this approach seeks to balance the social, psychological, and behavioural aspects of the illness with the biomedical perspective, which previously risked being undervalued. However, the first two stages still emphasise effective communication to build a good doctor–patient relationship. Therefore, to conduct a consultation with a patient, the doctor should carefully consider the consultation setting, perhaps by preparing a quiet consultation room and taking care about personal hygiene and professional dress. The interview should be conducted in an uninterrupted environment and started by establishing rapport with the patient, seeking to put him or her at ease with a feeling of respect and interest (Epstein et al., 2008).

The different consultation models mentioned above aim to lead the doctor to follow, defined, and explicit tasks to achieve an effective interaction. The Calgary–Cambridge approach includes an extensive list of micro-skills intended to make it suitable for managing a wide diversity of situations. Consultation models are learning aids used to help clinicians develop his/her own consultation skills, and they also help map out this aspect of clinical skills assessments.

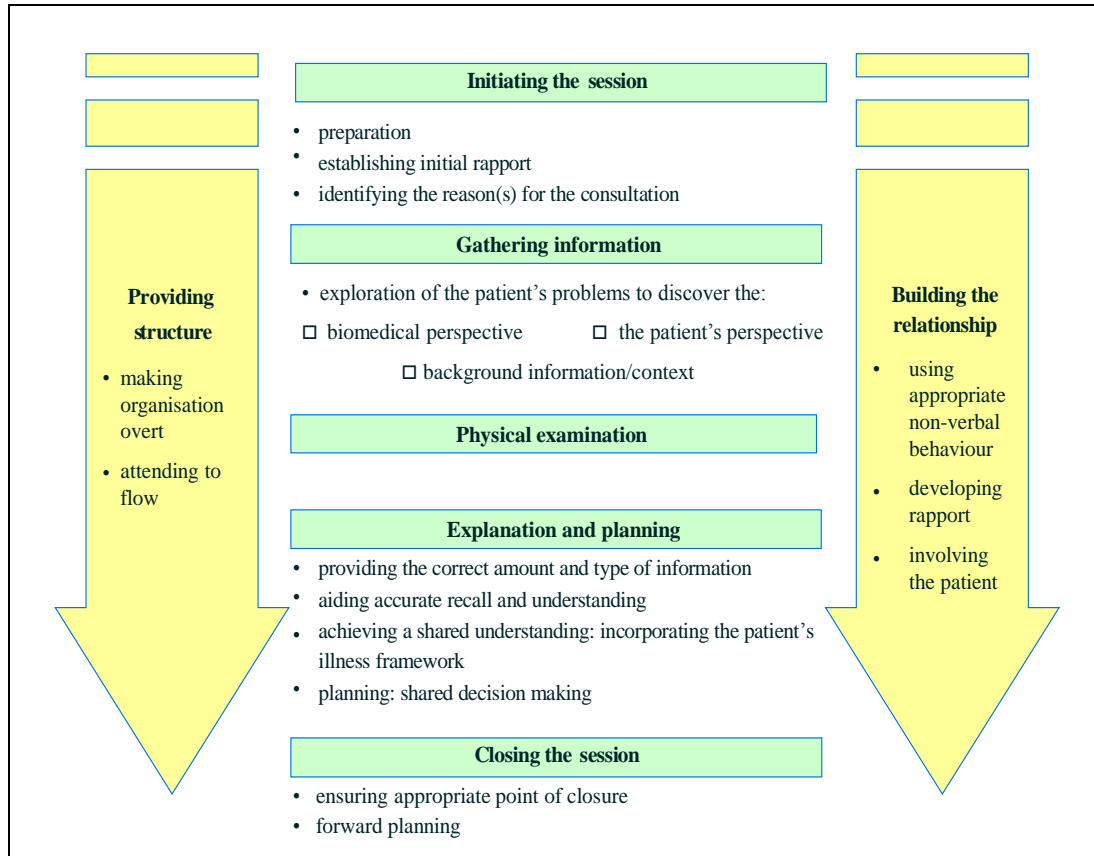


Figure 2.1: The Calgary–Cambridge approach in consultations.

2.4.2 Consultation skills and the doctor–patient relationship

Achieving an effective interview requires more than following an appropriate schema or model. The way doctors communicate with patients builds the doctor–patient relationship as much as what is said and in what order. Hence, many essential actions such as active listening, guided questioning, non-verbal affirmation, empathic responses, etc. as well as the ability to cover the necessary topics need to be mastered and appropriately employed (Bickley and Szilagyi, 2009). Doctors should be able to modify their behaviours to be more sensitive to patients’ physical and psychological components as required. Thus, specific training is needed to optimise doctor–patient consultation skills. In this regard, the importance of Balint’s work on raising the awareness of the importance of this relationship and the need for training becomes apparent (Balint, 1957); (Johnson, 2001).

2.4.3 The development of the doctor–patient relationship

‘Sir William Osler said, “The good physician treats the disease; the great physician treats the patient who has the disease.” The great physician understands the patient and the context of that patient’s illness. ’

(Centor, 2007)

Doctors and patients interact daily within various health settings, and this requires the creation of a ‘doctor–patient relationship’ which helps mould the behaviour of each with the other. This is now a recognised key element of western medicine. For instance, Helman described this relationship as ‘one of the most unique, but also problematic, forms of human relationship’ (Helman, 2003).

Though ‘good’ communication may be important, it can be difficult to provide consistently in daily health activities. Patients have concerns about their health and expect certain actions from physicians, including demonstrating awareness and acceptance of his/her feelings and values. Also, patients like physicians that are ‘present’ and are ‘with’ him/her (Spiro et al., 1996). These attributes seem of value for enhancing the doctor–patient relationship but are hard to define and achieve in face of other pressures, such as time.

A successful medical consultation depends on effective relationships between the doctor and his/her patient, and without this, the clinical knowledge or other skills of the doctor might be insufficient to reach an accurate diagnosis or discuss treatment options effectively, let alone engage patients in decisions about their care.

Historically, one of the early works on the doctor–patient relationship was reported by Balint in his book *The Doctor, His Patient and The Illness* (Balint, 1957). He proposed and developed the use of discussion groups to examine and build the doctor–patient relationship as a form of therapeutic intervention in itself. As a consequence, he significantly influenced the profile of training for the doctor–patient relationship for general practitioners (Johnson, 2001).

More recently, the importance of learning effective consultation skills has been included by the GMC as a key element of the foundation programme under the heading of ‘relation with patients’. The curriculum states (GMC, 2007):

‘Doctors must be able to develop, encourage and maintain
successful professional relationships with their patients.’

Patient-doctor interactions are better when doctors have effective communication abilities that enable them to obtain the information needed from patients to clarify their

problems. Training physicians in specific consultation skills can improve those skills and result in the reduction of patients' emotional distress (Roter et al., 1995). This leads to more accurate diagnosis, better care, and the identification of more emotional distress in patients (British Medical Association, 2011, Lloyd and Bor, 1996).

In the UK, the GMC consider a successful patient–doctor relationship an important element of professional behaviour for doctors or even medical students during their clinical training (GMC, 2009b, GMC, 2009c). Therefore, healthcare educators and providers seeking to improve health care and reach the highest standards of patient satisfaction should take the consultation or medical interview seriously. However, before investing in training, time and facilities, it is necessary to consider the supporting evidence that this is worthwhile.

2.4.4 The effect of the doctor–patient relationship

‘People don't care how much you know until they know how much you care.’

This quote has been attributed to the leadership expert, speaker, and author John C. Maxwell as a comment on people's behaviour. However, it can be modified slightly to use the word ‘patients’ instead of ‘people’, as follows:

‘Patients don't care how much you know until they know how much you care.’

This statement reflects patients' thinking towards clinicians. It reflects the high value of caring for somebody in addition to the more clinical aspects of the encounter.

The value of improving the communication between doctors and patients has studied by researchers through a number of trials. For example, Stewart reviewed 16 studies published from 1983–1993; concluding that patient health outcomes were improved (e.g. better blood pressure control, pain control, reduced anxiety, improved patient satisfaction, etc.) by an effective doctor–patient relationship, with a clear positive correlation between them (Stewart, 1995). Di Blasi examined 25 randomised controlled trials and concluded that the more effective doctors are those who are friendly to their patients and use a reassuring manner and informal environment during consultation (Di Blasi et al., 2001, Di Blasi Z et al., 2001). This study was followed by a systematic review conducted by Griffin (2004) to identify published randomised trials of interventions to alter the interaction between patients and practitioners. Based on 35 trials, Griffin opined that a range of approaches can achieve changes in interaction between doctors and patients.

Through their experiences with cancer patients, two main pathways (direct and indirect) through which doctor–patient communication influenced health and well-being have been identified. Communication can affect health directly; for example, the expression of empathy by a doctor may improve a patient’s psychological well-being (Schofield et al., 2003); (Ong et al., 2000). More commonly, communication can affect health through either proximal (e.g. satisfaction with care) or intermediate (e.g. self-management skills) outcomes (Stewart, 1995); (Epstein and Street, 2007). However either of these pathways (direct or indirect) may also depend on other factors (outcome of interest, the health condition, or simply the patient’s life circumstances); and caution should be taken when considering how this may contribute to health outcomes (Street et al., 2009). Therefore, there is some evidence that communication affects the clinical relationship and, thus, health outcomes. The

relationship between communication and health outcomes seems plausible and is now widely accepted, as reflected by the expectations placed on educators to include competence in communication at different levels of training.

2.4.5 The doctor–patient relationship in medical education

There have been significant changes in medical education over the recent decades and among them, an increasing emphasis on communication and the relationship between doctors and patients. Recently, the content for undergraduate clinical consultation curriculum was clarified by a consensus statement from all 33 UK medical schools (Fragstein et al., 2008). However, there are a range of approaches that may be taken.

The rapid growth of health technology, emerging patterns of chronic disease, rising health costs, changing social norms, etc. all have an effect on physician and patient behaviour. One of these effects has been a change in the role of doctors, with greater attention to the doctor–patient relationship and the provision of patient-centred care (Burke, 2008). This transformation in medicine has been reflected in medical education curricula and clinical training programmes. In 2003, the board of medical education of the British Medical Association initiated a discussion process that was subsequently published in 2004 and which emphasised consultation skills training based upon the evidence of improved health outcomes (British Medical Association, 2004). Since then, there has been broad acceptance of the idea that the way a doctor consults is important and that there are a number of different models available and an evolving literature on its teaching and assessment.

In 1991, Simpson made a clear and urgent call for teaching in doctor–patient communication and suggested the incorporation of this training into undergraduate and postgraduate medical curricula and continuing medical education courses

(Simpson et al., 1991). Others argued that this justified longer and intensified training periods, broader skills teaching, and the need for feedback on performance (Brown et al., 1999b). Attending training courses, Shilling argued, would increase doctors' insight into how effective communication improves the consultation and assists them to develop and utilise relevant skills in practice (Shilling et al., 2003).

Some studies found that physicians who participated in a training course on consultation skills demonstrated improved patient-centred behaviours and patient satisfaction (Fallowfield et al., 2002); (Jenkins and Fallowfield, 2002), although others failed to show this (Shilling et al., 2003); (Hulsman et al., 2002). This may be due to ineffective training intervention or the difficulty of detecting changes in communication behaviour which may be overshadowed by other factors, such as patients waiting too long to see a doctor. In addition, there were other effects on patients' satisfaction which might confound the results, such as age (i.e. older patients report significantly greater satisfaction with consultations than younger age groups) or psychological morbidity, (i.e. patients in distress report significantly lower scores than those with lower levels of psychological distress; (Shilling et al., 2003).

In 2004, the UK board of medical education published the following recommendation (British Medical Association, 2004):

‘Medical schools must be encouraged to provide the best possible communication skills training as part of the core undergraduate curriculum. This training should have set objectives and clear methods of assessment.’

However, another issue is the stage of training at which consultation skills training can be effective. The distinction between clinical and non-clinical years in medical

schools has diminished as more integrated curricula have evolved. These typically involve early contact between medical students and patients, providing opportunity for the early introduction of clinical consultation skills training. Higher levels of skills have been reported after the introduction of a new course in the pre-clinical stage of medical school (Evans et al., 1993b).

There has been concern that basic skills training in the early years may not suffice, and leading authors contend that teaching and training should require a comprehensive programme of increasing sophistication throughout medical school training and beyond (Kurtz, 2002). However, until 1998, only 13 out of 19 medical school in the UK assessed students' consultation skills, suggesting these skills were not considered to be of key importance (Hargie et al., 1998).

Medical schools aspire to reach the evolving international standard in medical education, and this may only be obtained by integrating consultation skills teaching with undergraduate curricula (Fragstein et al., 2008); (Tczac and Schwarz, 2000). Over the last 20 years or so, a consensus has emerged regarding the importance of and methods suitable for teaching clinical consultation skills. Attention is now focusing on the effective development and related assessment of appropriate levels and specific aspects of competence throughout training. For instance, a student in year one or two would not be learning skills at the same level as one in year five or a postgraduate. Likewise, an ability to apply an efficient and well-structured approach to 'taking a history' is not sufficient on its own, and learners must also be able to demonstrate what Kurtz and Silverman (1996) describe as 'perceptual' and 'process' skills. One of these is the ability to convey empathy.

However, if these skills are an essential requirement for medical education and practice, they also necessitate robust assessment (Epstein and Hundert, 2002). There is increasingly acknowledgement of the importance of the patient's perspective and of developing doctor–patient partnerships as a result of many studies that have linked empathy and consultation skills to patient satisfaction across a wide range of healthcare settings (paediatrics, stroke units, diabetic clinics, rehabilitation facilities, and eating disorder clinics (Wasserman et al., 1984); (Pound et al., 1998); (Hornsten et al., 2005); (McGilton et al., 2006); (Ramjan, 2004).

So there has been a major shift over recent years, and effective consultations are seen as a key element of healthcare which impacts directly on healthcare outcomes (both directly and indirectly), as well as patient satisfaction and perceptions regarding their care. There is also an ongoing shift towards greater patient participation in their care, service delivery, and medical education. In addition, it has become clear that empathy is the cornerstone of patient–doctor relationship as patients who feel fully understood by his/her physician feel respected and that they have great value (Beckman et al., 1994); (Levinson et al., 1997). The next section will focus on empathy and the empathic behaviour of health professionals during consultations, specifically.

2.5 Empathy in the healthcare context

In the previous section, we explored the role of the doctor–patient relationship as part of the consultation; here we will find to what extent empathy is crucial for this relationship. More recent research has suggested that empathy promotes both patient and doctor satisfaction and may achieve or improve the clinical outcome (Coulehan et al., 2001); (Hojat, 2007).

This section explores the definition of empathy and shows to what extent it specifically matters in the delivery of healthcare. It outlines concerns about an apparent decline in this among physicians, nurses, and medical students during their training, as well as support for developing empathic skills during medical education. Finally, the need to test or rate empathy is considered here, specifically in medical students.

2.5.1 Definition and historical background

The term empathy means ‘to share feelings with others and to understand what is beyond these feelings’. The verb ‘to empathise’ (with somebody) means to understand another person’s feelings and experiences, especially because you have been in a similar situation (Oxford English Dictionary, 2011). The origin of the word empathy appears to be in the year 1873, when the German art historian Robert Vischer used the word *Einfühlung*, which means ‘feeling into’ or ‘feeling within’ to describe the individual’s feelings when appreciating a work of art (Hunsdahl, 1967); (Jackson, 1992) (Wispé, 1986). The psychologist Lipps further developed the notion of *Einfühlung* (Lipps, 1903, Lipps, 1905) and formulated the concept of the ‘self-projected into the perceived object’. In 1909, the British psychologist Edward Bradford Titchener introduced the term *empathy* as an English synonym for the

German word *Einfühlung* (Titchener, 1909). Titchener translated Lipps notion of *Einfühlung* as *empathy* by way of the Greek *empathia*, meaning to appreciate the feelings of another person (Wispé, 1986).

The distinction between empathy and sympathy, specifically, has been debated since the time of the Scottish social philosopher Adam Smith (1723–1790; (Nightingale et al., 1991), and many psychologists have noted the confusion between the two concepts (Eisenberg and Lennon, 1983); (Olinick, 1984). Consequently, Wispé (1986) tried to clarify and rehabilitate the concepts of sympathy and empathy. He held the view that sympathy involves a sharing of emotions, which may be reflected in the expression ‘I feel sorry for you’. In contrast, empathy involves a sharing of understanding, as reflected in the expression ‘I understand how you feel’. A low degree of empathy might be noted by a client when he feels that his claim was not understood by the professional helper. In recent decades, researchers have struggled to find a clearer definition of empathy, having found it difficult to define and hard to measure (Kestenbaum et al., 1989). Therefore, nowadays definitions vary somewhat but have aligned around similar content (Hojat, 2007). Wispé (1996) defined empathy as ‘the attempt by one self-aware self to comprehend unjudgmentally the positive and negative experience of another self’. This was followed by another attempt by Carol Davis to distinguish the concept of empathy from other related terms because the confusion can lead to problems in medical education and patient care, where it may affect clinical behaviour and patient outcome (Davis, 1990). The difficulty with using the terms empathy and related terms interchangeably within literature and conversations was also noted by Davis (1990) and Nightingale et al. (1991), as it leads to an increase in confusion and makes the task of distinguishing empathy from these related terms more difficult (Davis, 1990); (Nightingale et al., 1991).

There have been a number of attempts to solve this confusion. The concept of empathy has been divided by the sociologists Mehrabian et al. and Davis into two aspects: vicarious and imaginative. Vicarious empathy is defined as an individual's vicarious emotional response to perceived emotional experiences of others and reflects an innate emotional response. The second type, imaginative empathy, is defined as an individual's ability to imaginatively take the role of another so as to understand and accurately predict that person's thoughts, feelings, and actions. It refers to a cognitive process and reflects a learned ability to imagine and intellectualise (Mehrabian A et al., 1988); (Davis, 1983b).

Hojat and his colleagues (2002a, 2003, 2007) have given careful attention to this and tried to clarify the ambiguity associated with definitions of empathy in the context of medical education and patient care as follows (Hojat et al., 2002b, Hojat et al., 2003, Hojat, 2007).

‘[Empathy is] ... a cognitive attribute that involves an understanding of the inner experiences and perspectives of the patient as a separate individual’

Furthermore, Hojat studied the definitions of empathy, reviewing discrepancies and concluding there is an overlap between them, which perhaps not surprisingly showed moderate correlation between measures of the two (Hojat, 2007, Hojat et al., 2011a, Hojat et al., 2001).

From the above, we can conclude that there is no one perfect or entirely agreed definition of empathy, which remains somewhat debated. Thus, this study has had to consider what aspect of the construct is most applicable in the clinical context as it

pertains to the medical profession, and has adopted the following definition from Coulehan and colleagues (Coulehan et al., 2001):

‘Empathy is the ability to understand the patient’s situation, perspective and feelings and to communicate that understanding to the patient.’

2.5.2 Empathy and quality of healthcare

Empathy is important for both physician and patient satisfaction and usually results from good communication between them (Wong and Lee, 2006). It may improve the quality of information gathered from patients and improve diagnostic accuracy (Coulehan et al., 2001); (Beckman et al., 1994); (Mangione et al., 2002). It has been found that patients feel more satisfied in the absence of physician domination and significantly improve when psychosocial topics are included as well as biomedical ones (Bertakis KD, 1991); (Kim et al., 2004); (Roter et al., 1997). Moreover, studies in dentistry affirmed the importance of empathy as communicative behaviour for dentists too (i.e. for reduction of patient anxiety during dental consultation (Corah NL, 1988); (Schouten et al., 2003). Furthermore, a cross-sectional survey showed a potential relation between empathy enhancement and increased resident well-being (Shanafelt et al., 2005). Davis argues the clinician must seek the client’s trust by the end of the consultation or the therapeutic process (Davis, 1990); yet, others have warned of the scarcity of empathy in modern medicine (Reynolds and Brian, 2000).

Neumann was concerned with empathy in clinical practice but faced difficulty searching for evidence to prove that it is a core element in the clinician–patient relationship. He warned about the lack of empirical research into empathy and

suggested a need for well-designed studies to overcome the complexity and multi-dimensionality of empathy (Neumann et al., 2009). In addition, Neumann developed the ‘effect model of empathic communication in the clinical encounter’, which demonstrated how an empathically communicating clinician can achieve improved patient outcomes. He analysed the nature and specific effectiveness of empathy and declared that the integration of clinician experiences with the situational factors seemed to be fundamental for development and promotion of empathy and to maintain high-quality patient care. In addition, the clinician can fulfil key medical tasks more accurately, thereby achieving enhanced patient health outcomes.

It has been affirmed that patients report more about their symptoms and concerns when they felt an empathic behaviour from their GPs (Squier, 1990); (Coulehan et al., 2001); (Neumann et al., 2008). This in turn leads to more accurate diagnosis (Larson and Yao, 2005); (Halpern, 2001), more illness-specific information given to patients (Kim et al., 2004); (Irving and Dickson, 2004), patients’ increased participation and education (Mercer et al., 2002b); (Kim et al., 2004); (Price et al., 2006), and even minimisation of patients’ emotional distress (Neumann et al., 2007).

The evidence for the patient outcomes described above is only linked indirectly with the physician’s empathy (Haslam, 2007). Instead, these positive outcomes have been associated with the patient–doctor relationship and communication practice, which has already shown to be enhanced by doctors’ empathy. However, direct links have been also reported, as doctors and medical students who are more empathic receive higher satisfaction ratings from patients (Colliver et al., 1998). In addition, it has been shown that medical students who are more empathic perform better in history taking and receive higher ratings for clinical competence (Hojat et al., 2002a); (Colliver et al., 1998). So it may be impossible to fully separate performance in terms of

empathy/caring from more tangible aspects such as history taking, as the effectiveness of the latter may itself influenced by improvement of the doctor–patient relationship.

Furthermore, in the context of therapeutic relationship; the American psychologist, Carl Roger, who developed the system of psychotherapy known as client-centred therapy, contended that empathy is an important key concept in the client-centred approach (Roger, 1951). Other psychologists have since explored the use of empathy and shown that an empathic listening approach was an effective specific component in a wide range of therapies (Peltier, 2001); (Bernstein and Nash, 2002); (Stober and Grant, 2006). Recently, a specific therapeutic effect resulting from physicians' empathy has even been linked to predicting the duration and severity of the common cold (Rakel et al., 2009).

2.5.3 The effect of empathy on patient enablement and health outcome

Research into the quality of general practice consultations in 1990s sought to move beyond satisfaction, and Howie and colleagues developed the concept of 'enablement'. They described patient enablement as (Howie et al., 1998, Howie et al., 1999):

‘the effect of the clinical encounter on the patients’ ability to
cope with and understand his or her illness.’

As a result, they developed the Patient Enablement Instrument (PEI) to assess the extent to which patients feel able to better understand and cope with his/her condition or health after their general practice consultation (Howie et al., 1997, Howie et al., 1999).

At the Glasgow Homeopathic Hospital, Mercer and colleagues validated PEI as a consultation outcome measure and demonstrated that doctors' empathy was considered a core determinant of patient enablement (Mercer et al., 2002b). Subsequently, MacPherson and colleagues used PEI and Mercer's research methodology in a retrospective study on patients at the Glasgow Homeopathic Hospital who had participated in a survey of acupuncture safety and found a significant positive correlation between the patient's perceptions of the practitioner's empathy with patient enablement (MacPherson et al., 2003). Similarly, Bikker and colleagues confirmed the strong relationship between consultant's empathy and patient enablement in their prospective study at the Glasgow Homeopathic Hospital. (Bikker et al., 2005).

The relationship between physician empathy and patient enablement at consultation has been confirmed in other primary care settings as well. In 2008, the results from a previous study comparing primary care in areas of high and low socio-economic deprivation (Mercer and Watt, 2007b) led Mercer and colleagues to conduct a pilot study exploring the relationship between GP empathy and patient enablement in an area of high socio-economic deprivation (Mercer et al., 2008a). Their results led them to conduct a larger study comparing this relationship in areas of high and low deprivation. This study affirmed that, although there are other factors that can influence patient's enablement, the need for physician empathy is mandatory for better patient enablement across different patient groups and conditions (Mercer et al., 2012). In addition, a link between physician empathy and patient enablement at consultation was also observed at a German oncology hospital (Neumann et al., 2007).

The relationship between clinician empathy and patient enablement is an important issue in the context of patient care. In their study, MacPherson et al. (2003) found an association between enablement and change in main complaint or health outcome but

could not confirm a direct correlation between empathy and health outcomes. However, other studies have demonstrated the importance of this relationship and the long-term impact of patients' enablement on health outcomes (Di Blasi Z et al., 2001). Regardless of whether these studies were on complementary (homeopathic or acupuncture) or conventional (primary- or secondary-care) therapies, they all had similar values for empathy mean scores using the CARE questionnaire (Bikker et al., 2005, Mercer et al., 2002b) or other measures of patient centredness (Little et al., 2001).

These results have been extended by other researchers in a variety of clinical conditions. For example, in a prospective study with acupuncture patients, perceptions of practitioner empathy at initial consultation and health benefits after two months were measured, and a significant direct association between empathy with patient' enablement (immediate outcome) and indirect one (prospective outcome) on health outcomes eight weeks later was found (Price et al., 2006). Furthermore, a recent study by Rakel and colleagues at the University of the Wisconsin, USA, suggested that empathy was useful in predicting the severity and duration of illness and immune response in patients with common cold (Rakel et al., 2009). Subsequently, in a randomised controlled trial of 719 patients with new cold onset, it was found that patient' perceptions of the clinical encounter were associated with reduced cold severity and duration (Rakel et al., 2011). These studies highlight the important role of the perception of empathy in the therapeutic encounter.

Thus, empathy and enablement are both measures of aspects of therapeutic relationship (Mercer and Reynolds, 2002a), and empathy appears required for patient enablement. Patient enablement plays a positive and important role in connecting empathy and

health outcomes, and the practitioner's empathy is essential for achieving better health outcomes via patient enablement (Price et al., 2006).

Recently, a systematic review was conducted by Lelorain et al. (2012) to investigate the association between empathy measures and patient outcomes in cancer care. The review revealed that clinicians' empathy has beneficial effects and related to higher patient satisfaction and lower distress especially with patient-reported measures. The clinicians' evaluations of patients' perspectives were highly recommended for future studies

2.5.4 Empathy in medical education

Empathy is a desirable ability that develops as individuals mature, and medical students are sufficiently cognitively and emotionally mature people who should have the ability to express empathy (Hojat, 2007). To some extent, these attributes (consultation skills, empathy, and other related personal attributes) are typically considered when selecting students (Albanese et al., 2003); (Lumsden et al., 2005); (Morrison, 2005); (Eva et al., 2004). Though this is done in general terms, no schools are known to set a minimum standard or measure empathy specifically well enough to apply a standard. So, more empathic individuals will probably have a better chance of gaining entry at schools, but many entrants may still have poor intuitive skills in this area and be required to improve these. It has also been reported that some medical students are concerned about their empathic skills (Tamburrino et al., 1993). Therefore, researchers have considered how to develop and promote these skills using different teaching approaches. On this subject, Stepien and Baernstein identified 13

peer-reviewed studies and opined that medical students' empathy can be promoted greatly through workshops in consultation skills (Stepien and Baernstein, 2006).

2.5.5 Empathy decline

Researchers using different methods and instruments to monitor the level of empathy in medical professionals over time have reported varied results. For example, when Diseker and Michielutte utilised Hogan's empathy scales with medical students of the class of 1979 at the Bowman Gray School of Medicine, they found that empathy scores (measured in 1975, 1976, and 1979) declined over time (Diseker and Michielutte, 1981). In 2002, Bellini and colleagues found similar results when they used Davis' Interpersonal Reactivity Index (IRI) instrument to measure empathy levels in resident doctors during the internship period: residents rated highly at the beginning of the internship, but scores decreased by the end of training period. Likewise, vicarious empathy was significantly decreased during medical education. (Bellini et al., 2002, Bellini and Shea, 2005).

Similarly, Hojat conducted an empirical study administering the JSPE instrument to medical students in different years at Jefferson Medical College, Philadelphia and found that empathy levels changed or declined during medical school (Hojat et al., 2004). Subsequently, Hojat conducted a longitudinal study (also using the JSPE instrument) on medical students. He found a significant decline in empathy during the third year of medical school when the curriculum shifts towards patient-care activities (Hojat et al., 2009).

However, other researchers have obtained different results. For example, Mangione reported that mean JSPE scores for internal medicine residents from the beginning to

end of the internship year did not change significantly (Mangione et al., 2002). Tavakol, using the student version (S-version) of the JSPE, found similar results for the fifth-year medical students at the University of Nottingham (Tavakol et al., 2011b), as did others using the IRI instrument (Zeldow and Daugherty, 1987); (Quince et al., 2011). Moreover, a cross-sectional study using a Portuguese adaptation of the JSPE showed that the empathy scores of Portuguese medical undergraduates in their final year were higher than those of students in their first year (Magalhães et al., 2011).

The debate about empathy decline has provoked two systematic reviews. In the first, Colliver and colleagues examined studies published between 2000 and 2008 investigating empathy in medical students and residents (Colliver et al., 2010). Their review of 11 articles using different rating scales revealed a small but significant decline in the mean ratings, although low response rates and other concerns about the validity of self-reported measures of empathy often used brought the validity of the finding into question. In the second review, Neumann and colleagues examined studies published from 1990 to 2010. This review reported that most of the studies showed a decrease in empathy during medical school and residency. The medical students' empathy declined steadily as they came close to clinical phase, where more contact with patients is required. The review also revealed that the majority of these studies (longitudinal and cross-sectional) were conducted in the United States, with only two of them conducted in the United Kingdom (Neumann et al., 2011a).

It is difficult to interpret these conflicting results; however, in general, recent studies that utilised self-administered instruments do appear to show a decline in empathy 'trend' with increased clinical training but an improvement in 'observed empathy', which might explain the conflicting findings. For example, while self-administered

instruments assess the internal emotion (trend), they may not predict how empathically the student behaves with patients (Chen et al., 2010a).

So perhaps both are correct, and therefore more sophisticated empirical longitudinal studies are needed to distinguish more clearly between how the clinician feels and how he/she acts (Magalhães et al., 2011). However, such analysis requires a more convincing instrument: one which measures the empathy perceived by patients, as opposed to the self-administered scales currently used for medical students and residents (Colliver et al., 2010).

Nonetheless, Bellini and Shea assert that there is considerable evidence of a decline in empathy levels during medical school and beyond and that this is directly proportional to the growing reliance on technology and the economic effect on health systems (Bellini and Shea, 2005). The abovementioned review by Neumann and colleagues (2011a) found that distress was a key factor in empathy decline when medical students and residents experienced burnout, suffering reduced quality of life or even depression (Neumann et al., 2011a). In addition, in a study on empathy in nurses, Baillie claimed that environmental factors such as workload and stress affected the nurses' ability to empathise (Baillie, 1996).

West and colleagues have contended that the cause of the decline is related to burnout, unhappiness, and loss of a sense of control during a long working day. They showed that medical trainers who experience a subsequent decline in empathy have increased major medical errors (West et al., 2006). In addition, the workload, long working hours, reduction in sleep may contribute to empathy decline in resident doctors (Rosen et al., 2006). In agreement with this, studies have shown a positive relationship

between well-being and empathy in students and doctors (Thomas et al., 2007); (Shanafelt et al., 2005).

Professional therapists and students must develop their empathy and not allow it to decline or become frozen. They must continuously improve this ability and practice it (Neumann et al., 2011a). Medical students and residents should constantly reinforce empathy in their consultation skills by the use of role-playing during undergraduate or residency training (Newton et al., 2008).

In summary, although the connection between an empathic nature or personality trait and consultation behaviours which may result is unclear, there is weak evidence that the former may fall during training and the latter may be affected by tiredness and workload as well as ability. This intriguing area is not the focus of this thesis but merits further investigation. Nonetheless, it is clear that research in this area has been hindered by the lack of a conceptually clear or perhaps even adequate definition of empathy and robust operational tools to measure it (Spencer, 2004).

2.5.6 The need for empathy measure

In spite of the apparent difficulty involved in developing a valid and acceptable measure of empathy, the theoretical import of the concept requires that continuing efforts be made.

(Hogan, 1969)

The changes evolving in the UK healthcare system and the expansion of market-driven managed-care delivery systems may hamper physician–patient relationships, and these

represent key drivers towards finding a viable way to measure empathy. Reynolds previously suggested that the lack of a good measuring tool for empathy might be considered one of the major reasons for low clinical empathy (Reynolds, 1994), and there is a risk this situation will deteriorate further.

Accordingly, many attempts have been made to develop a measure of empathic ability, particularly in psychiatry during the last century. Various measures have been developed which focused on the different aspects of empathy and were designed for use with the general population as well as those targeted specifically for health settings. However, the lack of a validated tool to measure empathy in patient-care situations has hampered researchers working with both clinicians and medical students (Hojat et al., 2003, Hojat et al., 2002a, Hojat et al., 2001). Also, few empathy measuring tools have been used in medical education research, with none content-specific to medical education (Hojat et al., 2009).

In 2002, Mangione and colleagues (2002) called for the development of an instrument to measure changes in empathy and asserted that this should be tested in large number of medical trainees and validated against both staff and patient rating at the same time. Such a measure, he argued, could be used to evaluate factors associated with empathy and the effectiveness of related clinical educational programmes.

‘In spite of the apparent difficulty involved in developing a valid and acceptable measure of empathy, the theoretical import of the concept requires that continuing efforts be made.’

(Hogan, 1969).

The measures developed for use in the general population have considered emotional, cognitive, and multidimensional aspects of empathy. Several measures have been developed to assess empathy in psychiatric or in nursing settings (in secondary care; (Burns and Auerbach, 1996); (Horvath and Symonds, 1991) and have failed to reflect the clients' views about the ability to offer empathy, perhaps because of the patient group involved.

In the context of medical education and general patient care, Hojat argued a clear definition of empathy could lead to content-specific measuring tools for empathy and thereby to enhance strategies to monitor and educate practitioners (Hojat et al., 2009). The section above has sought to explain the challenges involved in this task and the importance of achieving it. As it has been determined that a pragmatic definition of empathy is required, with the context anchored to undergraduate medical students undertaking generalist consultations, the next section will discuss the measurement issues that have to be considered when selecting and evaluating potential relevant tools.

2.6 Assessment in medical education

This section of my thesis focuses on the purpose of assessment, including its principles and general goals. The purpose of an assessment, whether summative (make a decision on outcomes), formative (provide feedback), or both is key to its design and needed characteristics. Summative and formative assessments will be reviewed to show the relationship between the two, and a description of the criteria required of methods in undergraduate medical education will be presented. The methods used in medical education need continually review to ensure that feedback to students helps build a strong background of consultation skills suitable for the doctor–patient relationship. This section also provides an overview of assessments of consultation skills in medical education and the main requirements for these assessments. Lastly, the section presents the importance of feedback in the consultation skills assessment process and its effects on learning.

2.6.1 General principles of assessment

It can be difficult to define assessment in an international context. The words ‘assessment’, ‘evaluation’, and ‘appraisal’ have similar meanings, and they are often used interchangeably. ‘Assessment’ is the term which refers to the judgment of students’ work in the UK and is similar to the term ‘evaluation’ used in the US. In this thesis, the term ‘assessment’ stands for any appraisal or judgment or evaluation of a student’s performance. However, we will also use the following definition of assessment:

‘the processes and instruments applied to measure the learner’s achievements, normally after they have worked through a learning programme of one sort or another’

(Wall, 2010).

Epstein summarised the principles of assessment as comprising four main domains: the goals of assessment, what to assess, how to assess, and cautions to be taken in assessment process. Keeping these principles in mind would enhance the current assessment practices with respect to the challenge of deciding whether to use formative or summative assessment formats, how frequently assessment should be made, and what standards should be in place (Epstein, 2007).

The assessment process is fully integrated in aspects of everyday life. However, in the clinical environment, assessment is a key component of learning. The assessment methods utilised in medical education are important for training and promoting doctors and students’ skills. The development of measurements of student performance that are both reliable have a formative educational role besides their value in clinical competence is considered one of the goals of assessment in medical education (Wass et al., 2001). In addition, it has been proposed that medical examining bodies should evaluate their assessment processes and look for appropriate methods to minimise sources of measurement error (Keen et al., 2003). Assessing students can provide opportunities to define individual strengths and identify areas of learning that need to be developed.

2.6.2 *Criteria for good assessment processes*

Assessment and feedback are key elements of undergraduate as well as postgraduate training and should be consistent with the education process. Assessment should be part of the learning process in order to achieve educational objectives set out in the training program (Van der Vleuten, 1996). Continuous assessment for new graduates is used for learning purposes, and rather than a test to pass or fail, it is an opportunity to learn and develop (GMC, 2011b). The package of assessment may differ from one student to another within the same medical school, as those who have difficulties need more extensive testing (Epstein, 2007). Generally, an assessment process aims to judge how much difference there is between the required standard and the goals achieved. This judgment could be made during or at the end of the assessment process.

The GMC's education committee set the standards for medical education in their successive editions of *Tomorrow's Doctors* (first edition in 1993, second edition in 2003, and third edition in 2009). A review of recommendation guidelines for these editions revealed that there is no a specific limitation or binding direction for medical institutions' methods of assessment for undergraduate students, as medical schools need to choose their own varied and justified methods of assessment (GMC, 2009a). The general guidance on assessment in the second edition of *Tomorrow's Doctors* stated the following:

‘Schemes of assessment must take into account best practice, support the curriculum, make sure the intended outcomes are assessed and reward performance appropriately.’

(GMC, 2003).

Based on this, the third edition of *Tomorrow's Doctors* has built a foundation for ensuring the validity, reliability, generalisability, feasibility, and fairness of the assessment process. In addition, it set criteria to determine whether the assessment process achieves the original specific outcomes (GMC, 2009a).

2.6.3 Assessment in undergraduate medical education

The assessment process is necessary to ensure that both the general public and regulatory authorities have confidence that the medical graduate can provide a required standard of healthcare and professionalism. There are different models of assessment in medical education. It is generally accepted that assessment drives students learning. As medical schools choose their own methods and standards of assessment, different assessment models can be found. This section will review the most models that are most related to the current study.

2.6.3.1 Miller's model of assessment

The framework for assessing clinical competence was proposed over 20 years ago by Miller in his famous pyramid (Figure 2.2) or model of clinical competence (Miller, 1990).

- 'knows': has adequate knowledge
- 'knows how': can apply that knowledge
- 'shows how': can demonstrate how to apply that knowledge
- 'does': applies that knowledge in practise

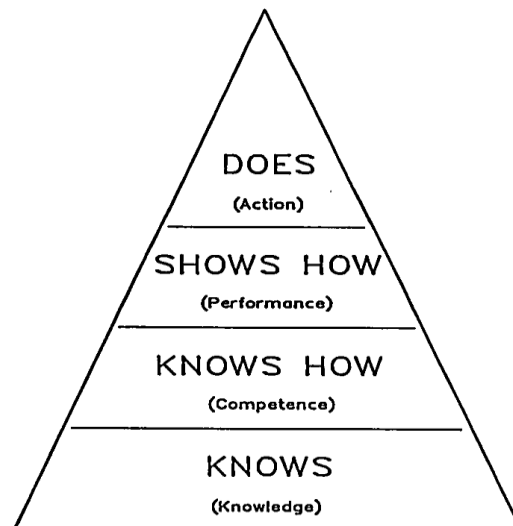


Figure 2.2: Miller's pyramid.

The lower two levels of this pyramid test knowledge or cognition. The upper two levels test behaviour. Using these four levels, it is possible to create an educational programme that begins with the collection of knowledge from books and articles and progresses through building clinical skills to reach real performance in practice. The 'knows' level is also called the 'recall level', where the student's knowledge is assessed. The 'knows how' level is the application of the recalled knowledge. The student's behaviour in skills and attitude is tested at the top two levels. Understanding these levels helps in choosing the best assessment tool.

Assessment at the 'does' level is more difficult than at the levels below, as it has to be observed through interaction with real patients, whereas at the 'shows how' level, simulated patients or demonstration of procedures could be used. The two highest levels of the pyramid represent the observation of clinical competence and performance. These observations vary from informal snapshot to complex and formal systems, which may include multiple raters providing assessment data about medical

students and trainers in different clinical settings during specified period of time (Downing and Haladyna, 2004).

Miller's model has some limitations. It does not explain this relationship clearly between competence ('can do') and performance ('does do'). Miller assumes that competence demonstrated by an examination will translate into actual workplace performance. In addition, he refers to the 'shows how' level as being an assessment of performance, although today, examination outcomes would be described as competency-based assessments. Moreover, his model fails to identify the influences of system-related and individual-related factors on clinical performance (Rethans et al., 2002). Therefore, Miller's model is no longer considered very useful in real practice assessment, as there are many contextual factors that influence clinical performance, such as time pressure, the mood of both the patient and doctor, or even the day of the week.

2.6.3.2 The Cambridge Model

In order to assess the doctor in practice, we need to distinct between competency-based and performance-based methods of assessment. Accordingly, Rethans et al. (2002) designated their Cambridge Model, which extends and refines Miller's pyramid (Figure 2.3). In this new model, performance is identified as a product of competence and highlights the influences of both individual (i.e. health, relationship) and system (i.e. facilities, practice time). In addition, it inverts Miller's pyramid and focuses mainly on the top two tiers.

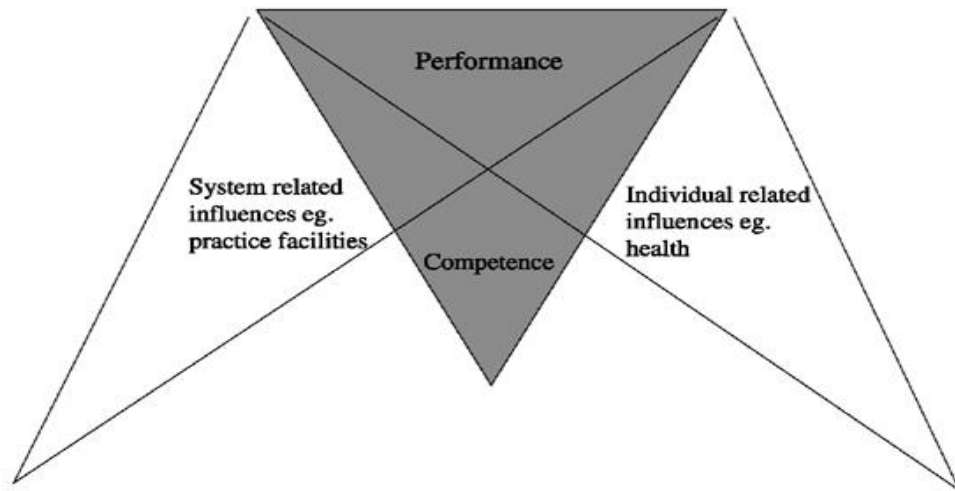


Figure 2.3: The Cambridge Model.

Assessment tools of clinical competence, such as the OSCE, are undertaken outside the real clinical environment and are assessed at the ‘shows how’ level of Miller’s pyramid. Other workplace-based assessment tools such as the mini-CEX and the Direct Observation of Procedural Skills are still largely ‘shows how’, as they are observational methods of assessment. The CARE questionnaire, which is given to patients in routine practice to measure patients’ experience of the empathy shown to them by healthcare providers, looks at the ‘does’ level of the pyramid and is a better example of an authentic ‘real practice’ assessment of the ‘does’ than the ‘shows how’.

2.6.4 Formative and summative assessments

There are two basic uses of assessment contained within two broad categories: *formative assessment* and *summative assessment*. Scriven (1967) was the first to coin and study these two terms. Each of these two approaches complements rather than contrasts with the other.

Formative assessment is assessment for learning and is also called 'educative assessment'. It is referred to as internal or 'Supervised Learning Events' and deals with how to make judgments about the worth of a programme while the activities are in progress. It is used to give feedback on checking the progress of the student and does not contribute to pass/fail decisions. This type of assessment helps both teachers and learners to find strengths and weaknesses and is best used when there is feedback from the student that reinforces the student's internal motivation to learn and improve their knowledge and skills (Wass et al., 2007).

Formative assessment is ideally performed several times during the teaching process, in order to improve, develop, and represent the judgments on that continuous process. It is accompanied by feedback to reduce the gap between the actual level of performance and the desired one and to find the ways in which the work can be improved to reach the required standard (Taras, 2005). In addition, formative assessment helps the educational process to adapt the teaching to meet the students' needs. It helps both teacher and student to monitor the progress and interfere if any modification in educational process is needed. Formative assessment is considered to give an opportunity for student to refine their thinking. Formative assessment functions as a benchmark for learning and motivates students to learn more and progress. It plays an important role in developing different aspects of medical professionalism and provides non-valued learning experience to students (Friedman Ben-David, 2000).

Summative assessment, or 'assessment of performance', is designed to inform a judgment on progress at a given point and can help determine the suitability of progression in training. It is a given end point during educational progression at which decisions may be drawn according to standards, goals, and criteria. It can form the basis of pass/fail decisions as to whether the student has reached the required standard.

The summative assessment is concerned with summing up the achievement of students and is reported at the end of a term or a course. Therefore, it does not provide immediate impact on learning but can be used for the purpose of certification (Sadler, 1989). This form of assessment was traditionally used to focus on outcomes and, according to Scriven (1967), all recorded assessments can be summative (e.g. have the potential to serve a summative function), but only some have the additional capability of serving formative functions.

The recent views towards summative assessment considered this type of assessment as a method through which 1) both medical graduates and future patients could be protected by denying graduation to students shown to be incompetent for patient care (Albanese, 1999), 2) students' awareness towards their behaviour could be increased at earlier stages of training (Crossley et al., 2002a), and 3) medical staff could be motivated and their clinical performance improved (De Haes et al., 2005), all of which would help to increase public confidence in the medical profession (Crossley et al., 2002a); (Epstein and Hundert, 2002). The GMC assessment strategy considered the summative assessment as a formalised assessment on which decisions about progression would be made, while the formative assessment was considered to be informal, frequent, dynamic and non-judgmental (GMC, 2011a).

2.6.4.1 Combining formative and summative assessments

The relationship between formative and summative assessments has concerned researchers. Careful attention on using these important elements of assessment should be taken when planning clinical competence testing (Wass et al., 2001). The traditional views towards the assessment process were stated by Scriven (1967), who claimed that

it was a single process and that there was a very fine line between formative and summative assessment. In addition, he believed that the summative assessment process stopped at the point of judgment and that if it were followed by feedback, it became formative. Nowadays, educators face many challenges in the assessment process, such as whether to use the summative or formative assessment format and how to determine the frequency of assessments, in addition to the challenges of keeping up with newly developed tools to assess professional and expertise skills (Epstein, 2007).

Epstein (2007) claimed that even without feedback, the summative assessment can influence learning process of students who are studying what they expected in the exam. In contrast, Taras (2005) suggested that formative assessment could focus on either process and product or both. Therefore, he preferred formative to summative assessment because formative assessment encompasses and justifies the summative one.

Summative and formative assessments complement each other and have been described as two sides of the same coin. Van der Vleuten and Schuwirth (2005) contended that both types encourage learning with very slight difference between them and that they can both be used in comprehensive assessment programmes. In addition, a recent approach to using combined formative and summative aspects in assessment of professional behaviour was suggested by Van Mook and colleagues (2010), who claimed that both summative and formative aspects did not contrast with each other but instead complemented each other, and presented the Maastricht medical school framework for professional behaviour as a real-life example of combining these two approaches, specifically when feedback is the core part of assessment process (Van Mook et al., 2010).

2.6.5 The importance of feedback on consultation skills

Feedback is considered a key element of formative assessment. It can be defined in terms of information about how successfully something has been or is being done (Sadler, 1989). Ramaprasad (Ramaprasad, 1983) defined it in term of its effect as follows:

‘Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way.’

Feedback is a core element of the assessment process, provided that it is appropriate, given in time and delivered in an acceptable manner (Morrison, 2003). Assessment methods that can provide valuable feedback for students and educators must help students to engage with their feedback and help them develop self-directed and independent learning skills to take charge of their own learning. Although the foregoing has been one of the most important goals of *Tomorrow's Doctors* since 2003, due to a shortage of staff, approximately 50% of UK medical schools did not comply with Paragraph 85, which states (GMC, 2011a):

‘Students will have regular feedback on their performance’

A review conducted by Sadler (1989) of textbooks on assessment revealed that very few textbooks published before 1989 (with a few notable exceptions) mentioned feedback and formative assessment in their body texts or index, although he opined that formative feedback improves learning efficiently and appropriately. In addition,

feedback has positive effects on both teachers and students. The teachers may make programmatic decisions with respect to readiness, diagnosis, and remediation, and feedback can be used by students to monitor their performance in order to reinforce the successful aspects or improve the unsatisfactory aspects (Sadler, 1989).

Likewise, feedback on student performance and discussions with tutors have a considerable effect on learning consultation skills (Maguire et al., 1986). Assessment methods that lack feedback may leave students confused about which are their weak or strong areas with respect to the subject matter, and, consequently, they may struggle to develop their learning (Epstein, 2007). It was suggested in *Tomorrow's Doctors* (GMC, 2009a) that patients and health professionals who come into contact with students should have an opportunity to provide constructive feedback. This feedback could help to identify strengths and weakness in both students and the curriculum.

Patients' views (feedback) started to be used to improve health services in the 1980s (Baker, 1990). Two decades later, Evans and colleagues (2007) called for the continuous measuring of patients' experience of healthcare and for further research into its provision and utility, given that patients are the end-users of healthcare and their feedback should be of pivotal importance to quality improvement.

2.6.6 *Methods of feedback in medical education*

Models for feedback have developed rapidly in the last two decades. The traditional model of verbal feedback does not take into account the complexity of feedback in learning (Kluger and Van Dijk, 2010). Ideally, feedback should provide an opportunity for trainees to improve their performance rather than threaten the progress as well as being essential for professional development (McKimm et al., 2009). The 'feedback

sandwich' (Docheff, 1990) is one of the best-known published models in medical education. It was designed to protect the learners' self-esteem and motivation, as you first make positive statements, then discuss areas for improvement, and then finish with more positive statements. Thus, in this model, negative comments are sandwiched between two pieces of positive message (Molloy, 2010).

David Boud and Elizabeth Molloy (2013) defined important key issues associated with feedback and discussed different models of feedback that aim to make feedback useful, relevant, and supportive to students. They looked at the whole feedback process from a student-oriented perspective and pushed towards not considering feedback as separate action from the whole module design process. They defined feedback as

‘a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, in order to generate improvement.’

Boud and Molloy reached that definition of feedback after they examined the commonplace assumptions about feedback: 1) all feedback is good, 2) the more feedback the better, and 3) feedback consists of telling people how to do something better, and once they have been told, that is the end of the story. The issue of students' feedback in higher education is a confusing matter which led both of these authors to develop a new way of thinking about feedback by analysing two main models of feedback in a recent study (Boud and Molloy, 2012). In the first model, feedback as driven by teacher to student, as it is in the disciplines of engineering and biology. In

the second model, students played a key role in driving learning by generating and advocating their own feedback and being the judges of the learning process. The goal of this second approach is to put the student in a proactive rather than a reactive position with respect to the process of feedback (Nicol and Macfarlane-Dick, 2006). In a similar manner, Orsmond and colleagues (2013) presented new methods of feedback delivery as an alternative to the old, standard methods, as shown in Table 2.1.

Table 2.1: New feedback delivery compared to the standard feedback methods.

New feedback delivery	Standard feedback delivery
Encourages dialogue between giver and receiver of feedback	Monologue; often tutor directed; one-way feedback
Involves peers	Does not involve peers
Explicitly encourages self-assessment/Regulation	Does not explicitly encourage self-assessment/regulation
Feedback on assignment process	Feedback on assignment product
Students encouraged to be <i>proactive</i> in working with feedback	Students encouraged to be <i>reactive</i> in working with feedback

2.6.7 Assessment as programme

Medical schools can improve their assessment systems through regular review of their reliability, validity, and educational impact of used instruments. The recent views about assessment have moved towards thinking of assessment as a programme in line with the broad education process required of a doctor and not possible to be measured by any single ‘best’ instrument. In their study, Schuwirth and Van der Vleuten (2004a) reviewed three studies that demonstrated these new views about the quality of assessments. The first study focused on the nature of the instrument examined and its uses (Basu et al., 2004). The second one focused on the effects of in-training assessment on students’ perceived level of confidence (Ringsted et al., 2004). The third studied the overall usefulness of a complete assessment programme (Daelmans et al., 2004). These studies moved the old thinking of using a single assessment tool towards seeing assessment as a programme underpinned by a suite of tools and methods.

Furthermore, using several different instruments with multiple observations can compensate for any defects present in a single instrument over time (Wass et al., 2001). Other studies have also considered the use of multiple methods of assessment to overcome the limitations of using a single measure format (Epstein et al., 2004); (Van der Vleuten, 1996). For example, Epstein and colleagues (2004) utilised the comprehensive assessment method to assess the habits of competence in the new curriculum of the second year medical students at the University of Rochester. They created exercises that showed the integration of different areas of competence by many previously validated formats (i.e. multiple-choice questions) with newer formats (i.e. peer assessment). They found that the comprehensive model of competence was feasible, useful to students, and, in the first instance, reliable and valid.

Although there is no perfect assessment tool, the one chosen should be specific to the knowledge, skills, and attributes being assessed. In addition, these tools should be feasible and consistent with the educational principles of the curriculum as well as being reliable and valid (Shumway and Harden, 2003).

2.6.8 Competency and outcome-based assessment

In the late twentieth century, medical schools in Scotland started to move their curricula away from the traditional approach towards learning outcomes. Accordingly, similar changes in their assessment system have been encouraged to ensure that students cover the desired learning outcomes set by the new curricula. For instance, in Dundee, a new outcome model, which relies on the three-circle model for classifying learning outcomes, was adopted. Figure 2.4 shows the essential elements of competence represented as three dimensions of doctors' work (Harden et al., 1999):

- 1) The performance of tasks: is the doctor is able to do the right thing?
- 2) The approach to tasks: is the correct approach to the task taken?
- 3) Professionalism: is he/she is the right person to perform the task?

Accordingly, 12 learning outcomes have been identified for Dundee curriculum to cover these three elements. This outcome learning model has been adopted by all the Scottish medical schools and subsequently implemented in their undergraduate curricula.

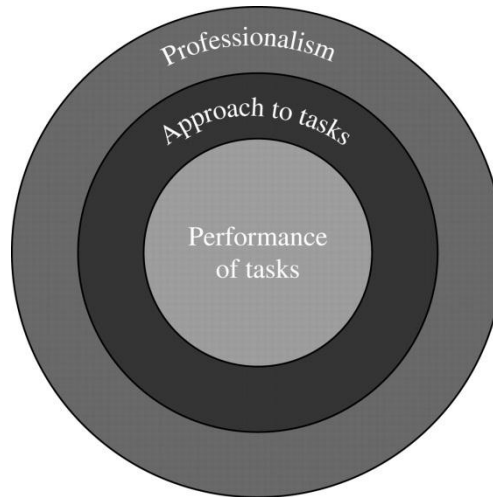


Figure 2.4: The three-circle model for outcome-based education.

Traditional curricular courses specified a defined duration but did not specify what skills or knowledge the student must acquire or what grade would be awarded at end of the course (Harden, 2007). In contrast, the availability of well-defined and validated competencies is very important for competency-based curricular models but needs to be applied appropriately to ensure educational standards. In addition, competency-based models place a greater burden on administrators charged with running the educational programme (Leung, 2002).

Although the outcome-based approach is not without concerns, it provides medical educators with a powerful instrument for assessment and helps both student and teacher in their mission for better educational delivery. Competency-based assessment programmes should be clear and have defined measurable learning outcomes that students can achieve at the end of the programme (Harden, 2002). Learning outcomes have been viewed as an important target for assessment, and the assessment methods used in medical schools curricula should adequately match these outcomes (Schiekirk et al., 2012).

Likewise, the assessment programmes in the medical schools need to meet the GMC requirements specified in *Tomorrow's Doctors* (GMC, 2009a). Learning outcomes cannot be adequately assessed by any single measure format if the required knowledge, skills, and attitudes are to be covered. Multiple methods are needed. This drive for clear and well-defined learning outcomes which must be achieved before the end of any course are well recognised as an important characteristic of outcome-based education (The Scottish Doctor, 2007).

2.6.8.1 Blueprinting

Blueprinting is the process of planning test content against the learning objectives of the curriculum. It is an important powerful tool for an integrated curriculum as it maintains assessing all its intended learning objectives (Hamdy, 2006). The test blueprint is used to cover the learning objectives and assess methods to identify the key topics which must be tested.

The process of blueprinting is simpler in the case of undergraduate assessments than for postgraduate assessments, as the core content is well defined with the first group (Wass et al., 2001). In the case of medical undergraduate programmes, the future responsibilities of the job help determine the competencies that should be acquired by the end of the programme and thus helps to build the assessments suitable for each level of training. The process of identification and analysis has an important impact on the educational process. It allows educators to plan and critically review their curriculum contents and learning strategies, and ultimately, to revise student assessment programmes (Hamdy, 2006).

2.7 Key concepts in assessment

This section reviews the concept of the utility of an assessment and places focus, in particular, on the reliability of the different forms of validity of exhibited by assessment tools in medicine. This section will also outline the steps needed to improve the reliability and validity of instruments used in medical assessment. In addition, the generalisability theory (G-theory) and decision studies (D-studies), which are statistical models used in psychometrics measures, will be covered.

2.7.1 *Utility of assessment*

The utility of the assessment system should be considered if used for learning or training. The utility of any assessment (meaning its usefulness) was defined by Van der Vleuten (1996) as a product of its reliability, validity, cost effectiveness, acceptability/practicality, and educational impact. Designing the assessment process should take into consideration all the elements within the following utility equation:

$$\text{Utility} = \text{educational impact} \times \text{reliability} \times \text{validity} \times \text{cost} \\ \text{effectiveness} \times \text{acceptability/practicality}$$

In addition, it should consider *feasibility*, which was added later on to the above elements because of its importance, although it is implicit in cost effectiveness and acceptability.

The components of the utility index are all important, and we should balance them according to the purpose and nature of the assessment programme or tool. For

example, reliability and validity may be more of a focal point in a high-stakes examination than educational impact, whereas the case might be reversed if we seek feedback from the trainee (PMETB, 2007). The utility model also helps educators make considered choices in selecting, constructing, and applying an assessment instrument. In addition, it helps in the process of designing an overall assessment programme for a whole course (Van der Vleuten and Schuwirth, 2005).

Schuwirth and Van der Vleuten (2004b) strongly believed that assessment instruments are not goals in themselves but that the first three criteria (validity, reliability, and educational impact) in the above equation are meaningful for the future of the assessment. This section will try to shed more light on the immanent characteristics of any assessment instrument according to the utility model, with a focus on the validity and reliability of the measures.

2.7.2 The validity of the assessment tools

Validity is defined as the extent to which a test measures what it is supposed to measure, and a valid instrument is one that which achieves the intended standard of measure (Norman et al., 2002). In addition, validity often defined as the extent to which an instrument measures what it purposed to measure (Kimberlin and Winterstein, 2008). In other words, validity describes how well one can legitimately trust the results of a test as it developed for a specific purpose.

Attention has been paid to validity of psychometric examinations conducted during the 1990s in conjunction with the introduction of the new undergraduate medical curricula (GMC, 2010b). Medical educators always have been concerned about the right task

being assessed in the right way by the right assessors, and whether the assessment is having a positive effect on learning and professional behaviour.

Generally, evidence of validity is required for all evaluations in medical education in order to be meaningfully interpreted (Downing, 2003). At the same time, multiple sources of evidence may be essential for evaluating the validity of any assessment. Validity is a feature of inference: there is no valid or invalid instrument, and there are no perfect instrument scores; but there are more or less accurate scores (Cook and Beckman, 2006). There are three important components that need to be considered that may interfere with validity:

- 1) the measurement tool (the form of the test or the type of the measure, e.g. observational, self-report, interview, etc.);
- 2) the purpose of the test (what we want to measure); and
- 3) the population context (for whom the measure is intended).

Caution should be taken when we use the term validity to avoid confusing the different forms of the term. The three traditional main headings of validity are content validity (including curricular validity and face validity), criterion-oriented validity (including concurrent validity and predictive validity, depending on the time we collect the criterion data), and construct validity (although contemporary thinking suggests that all validity should be conceptualised under 'construct validity'; (Cook and Beckman, 2006).

Content validity ensures 1) that there is a statistically significant agreement between experts in the questions asked and 2) that the instrument assesses the desired construct. It is the degree to which the measure is agreed as sampling the subject in question. For

example, the content validity of a test could ensure that the learning objectives of the course are covered by the questions assessed, to ensure that the test is appropriate for the programme's educational objectives and the target group assessed (Norman et al., 2002). Content validity should be a prominent feature of any assessment method and must be designed and mapped on a blueprint (Van der Vleuten et al., 1991).

Curricular validity, while similar to content validity, considers the quality of the whole course to meet its intended objectives. Curricular validity depends on the judgment of experts in the field. (Kimberlin and Winterstein, 2008). Face validity (or acceptability), a weaker form of content validity, is usually used to indicate the validity of tests at first impression (Van der Vleuten, 1996) and to describe the appearance of validity in the absence of empirical testing; in other words, does the test seem to be reasonable? Face validity is a judgment based on subjective perceptions of an instrument (Cook and Beckman, 2006) and is one of the key factors affecting the utility of an assessment according to the Van der Vleuten conceptual model (Van der Vleuten, 1996); (Van der Vleuten and Schuwirth, 2005).

Criterion-oriented validity is subdivided into concurrent validity and predictive validity. Concurrent validity is where a measure demonstrates similar results to existing measures of the qualities tested. It refers to the extent to which two (or more) tests of the same construct correlate. For example, does a new improved test find results that are similar to those of an existing test? If so, then this is encouraging and the new test shows concurrent validity. If, however, there was no correlation, then the tests are testing different constructs and are not related measures. Alternatively, if the new test correlates perfectly with the old test, then there is no point in instituting the new test on the grounds of validity alone (Streiner and Norman, 2008).

Predictive validity is the instrument's ability to show or predict an outcome, for example the success or failure of training outcomes. This type of validity can also predict an outcome in the future from given performance (e.g. professional success after graduation). This is clearly of potentially high value to the design of a training programme.

Construct validity is the extent to which the measuring tools succeed in assessing the construct for which it is designed. This is achieved when the instrument demonstrates and supports an outcome hypothesis (Norman et al., 2002). For example, experts perform better than novices on the test. This type of validity is based on the accumulation of evidence from previous studies that used a certain measuring instrument (Kimberlin and Winterstein, 2008).

Five sources of evidence to support construct validity have been described by Messick: content, response process, internal structure, relations to other variables, and consequences. These cannot be considered as types of validity but rather are categories of evidence that support construct validity (Messick, 1989). Similarly, Downing (2003) concluded that reliability is a major source of validity evidence for assessments. Convergent validity is the degree of correlation that a measure theoretically predicted to correlate with other measures of the same trait has. Conversely, discriminant validity refers to low correlations when the same method is applied to different traits.

2.7.3 The reliability of assessment tools

The reliability of an assessment is 'its ability to show stability and consistency over varying conditions of measurement (e.g. different observers, physicians, patients)'

(Norman et al., 2002). Here, reliability means the ‘consistency’ or ‘repeatability’ of the research measures. The origin of the definition of reliability is still unknown. Although it was mentioned for the first time in a textbook published in the 1930s, it is difficult to achieve a real understanding of the reliability concept; however, it usually reflects how much error is inherent in any measurement. Therefore, the reliability of the measuring tool is determined by the quality or accuracy of a measure.

The first measure of reliability was described by Fisher in his book on statistics. Fisher called this measure the ‘intraclass correlation’ (multiple observations of the same variable), to differentiate it from Pearson’s correlation, which is known as ‘interclass correlation’ (multiple observations of different variables) (Fisher, 1925).

Reliability is often reported as a coefficient called Cronbach’s alpha (Cronbach, 1951; see Section 2.7.4), which is determined using the following equation:

$$\text{reliability} = \frac{\text{subject variability}}{\text{subject variability} + \text{measurement error}}$$

However, reliability cannot be calculated accurately; it can only be estimated. This is because nobody knows the true score of specific observation (Trochim, 2006). Van der Vleuten contended that the reliability of a test can be improved by using a larger sample size and number of raters, examiners, or patients (Van der Vleuten, 1996). In addition, unstandardised and non-generalised assessment methods are considered to be unreliable (Streiner, 1985); (Van der Vleuten, 1996).

For repeated measures, analysis of variance (ANOVA) can be used to estimate reliability when different variance components have been merged in the coefficient. Any subject (e.g. student) may be repeatedly observed by different observers (examiners) many times using different kinds of assessment tools. An ANOVA table

then could be constructed by finding the sum of squares for the three variants (observer, subject, and error). Random error is contributed by the mean squares of the subjects and observers, and this leads to the variability of the scores. Therefore, at the end we can obtain the reliability coefficient from the variance between subjects and the error variance.

Reliability coefficients range from 0.00 to 1.00, with a higher coefficient indicating a higher level of reliability. In case of a high-stakes medical examination, a coefficient of approximately 0.9 is the target; any component with reliability below 0.8 is considered inadequate (GMC, 2010b). When there is no error in a measurement, this means that the scale is very reliable (Streiner and Norman, 2008). Thus, the smaller the error variance, the higher the reliability coefficient.

2.7.4 Types of reliability

Reliability can be considered by considering the potential for error from different sources (e.g. question, raters, and times of measurement). By understanding the error contribution from each potential source in an assessment, the best and most feasible blend of observations can be recommended for any given test to meet the required level of reliability for a test's purpose. There are four classes that will be considered in the present study: inter-rater or inter-observer reliability, test-retest reliability, parallel-forms reliability, and internal consistency).

Inter-rater reliability is represented by a number from 0 to 1 and gives the confidence by which an additional rater score might be predicted by the opinion or average of scores of existed raters. There are two methods to estimate inter-rater reliability. First, one simply calculates the percentage of agreement among the observations of raters.

This will give an idea whether agreement exists; but at the same time, it is a crude measure. In the case of continuous measurement, the estimation of reliability can be established through calculating the correlation among the observations of different raters. Although this method of estimating reliability requires multiple raters or observers, it is still considered as one of the best ways to estimate reliability when the measure is by observation.

Test-retest reliability is represented by a number from 0 to 1 and gives the confidence by which results from an additional time of measurement might be predicted by the opinion or average of opinions of existing occasions of measurement. Test-retest is also called the stability of the measurement. To estimate test-retest reliability, the same test should be applied to the same sample on two separate occasions. The test-retest estimate of reliability can achieve considerably different estimates depending on the interval. The time factor is extremely valuable in this type. The smaller the time gap, the higher the correlation, because of the relationship between two observations over time.

Another form of reliability, the internal consistency of an instrument, depends both on the similarity in the results obtained from different items or questions asked by an inventory and the extent to which these items are consistent with one another in a same measure (i.e. items measuring the same construct should correlate). Accordingly, the reliability among different items can be calculated by using correlations between each pair of items while measuring the same subject.

The value of internal consistency is represented by Cronbach's alpha (Cronbach, 1951), which is widely adopted as a measure of the reliability (internal consistency) of psychometric instruments (e.g. questionnaires; (Bland and Altman, 1997); (DeVellis,

2006). Cronbach's alpha represents the average inter-correlations of items and the number of items in the scale. Multiple items help in the determination of reliability of measurement and in general improve the precision of the measurement. Therefore, the higher the number of items in a scale, the larger Cronbach's alpha tends to be (Kimberlin and Winterstein, 2008).

G-theory is a mathematical technique that was originally developed by Cronbach and colleagues (1972) to address multiple possible sources of error variance in a measure. These sources of error are termed 'facets', and G-theory analyses the variance of each facet. Thus, facets contributing to random variance (e.g., time, questions, and times of measurement) can be identified and then minimised (e.g. by using more raters, questions, or times of administration) and consequently, the subject of discrimination (e.g. student performance) can be maximised.

G-theory also allows the collection of data from a study to be modelled mathematically using a technique known as a D-study. (Streiner and Norman, 2008). In this way, the best blend of resources (examiners, number of questions, and times of assessment) can be calculated. G-theory measures the reliability of obtained subject scores by quantifying the possible sources of rater and item variance (Streiner and Norman, 2008) (Crossley et al., 2002a). The aims of G-theory, to identify the discriminating ability of the test and integrate this with the reproducibility of the result, depend on identifying and removing any foreign or incidental variance which may affect the true scores of the main subjects of that measure (Narayanan et al., 2010). By generalisability analysis we can identify the sources of unreliability of the test whether in its items or observers (Thorndike, 1982).

The source of difference or the variance (facets) can be found in subjects, assessors, times, task/cases, and questions. The number of each one of the mentioned variance is concerning researchers. We should decide who we want to differentiate (diff) between. Also, we need to know what facets we should focus on and wish to generalise our findings (for example, questions, times of measurement or raters) (Table 2.2). Depending on our facet of generalisation, we may be able to take out (fix) the effect on reliability of another facet of the measure.

Table 2.2: The types of reliability with the sources of differences (facets).

Type of reliability	Facet of differentiation	Facet of generalisation	Facet fixed
Internal consistency	Student (Subject)	Question	Rater, Time
Inter-rater reliability	Student (Subject)	Rater	Question, Time
Test-retest reliability	Student (Subject)	Time	Rater, Question

The following formula is used to calculate generalisability:

$$G = \frac{\text{Variance (V) (Diff) + V (Diff X Fixed)}}{\text{V (Diff) + V (Diff X Fixed) + V (Gen X Fixed) + V (Diff X Fixed X Gen)}}$$

The G-study outcome is called the G-coefficient and provides researchers with confidence that they will obtain similar results when they repeat their test in different environment, raters, and subjects. Crossley considered that a G-coefficient = 0.80 is an acceptable level of reliability in professional assessment studies (Crossley et al., 2002a).

D-studies are used to determine the effect of changing the number of observations and to handle predictions of minimum numbers of raters or items required to generalise the results of a certain study (Streiner and Norman, 2008). For example, if a study has a G-coefficient of 0.90 and there are fixed number of raters per subjects, a D-study may show that we can reduce the number of those raters per subject and still achieve a respectable G-coefficient at a level of 0.80. D-studies depend on balanced design (i.e. same number of raters for each subject) and it is not clear how to amend and use a D-study with unbalanced studies (Narayanan et al., 2010).

2.7.5 Reducing measurement error and improving reliability

Reducing error in the measurement process should be of great focus when we develop and validate an instrument. It is the responsibility of the test developer to identify the sources of measurement error (Crocker and Algina, 1986). Unreliable assessment, whether the assessment is summative (used for decisions on outcomes of assessment, e.g. pass/fail) or formative (feedback; to promote learning), is of little help to examiners. Also, it cannot help medical schools in determining strengths and weakness of graduates' skills (Norcini, 2000).

It is the responsibility of the researchers and educators to refine their instrument and to identify the sources of measurement error by pretesting or pilot testing (Crocker and Algina, 1986). The development of instrument reliability concentrates mainly on reducing error in the measurement process (Kimberlin and Winterstein, 2008). In medical education, as well as other subjects, there is no absolutely reliable exam, and it is difficult to improve reliability because of two potential difficulties:

- 1) All assessments have a degree of error. This may be reduced but cannot be removed completely.
- 2) Difficulty arises if there is a uniformly high level of achievement and thus a limited range of performance. Medical students' assessments sometimes show high marks with little variability, and this limits the capacity to discriminate performance and so show reliability of the assessment.

To improve reliability, we have to reduce the error effect and design the assessment in a way that differentiates between candidates. Reducing measurement error can be considered during instrument development, data collection, creation of the dataset for analysis, and collecting feedback from the respondents followed by analysing their comments. We might ask whether the instrument was easy or hard to use. In the case of gathering measures, appropriate supervision and training of data collectors must be conducted carefully, so that they do not inadvertently introduce errors (Trochim, 2006); (White et al., 2009). In addition, the data for analysis should be double entered, so that it can be verified. In addition, study results can be adjusted for the effect of measurement error by using information from a validity or reliability study.

Researchers should identify sources of measurement error by pretesting or piloting the test instrument and refine that instrument to minimise the potential error (Crocker and Algina, 1986). An important method of minimising measurement error is using multiple measures of exposure for the same subject (e.g. mean, standard deviation). It is widely used to deal with systematic measuring errors, especially when the different measures do not share the same systematic error (White et al., 2009).

2.7.6 The relationship between reliability and validity of assessment tools

The reliability and validity of a measure are interrelated. In fact, the maximum possible validity of a test is dependent on the square root of its reliability (Streiner and Norman, 2008). This relationship is intuitive; that is, one would not wish to base a decision on a measurement tool that gave a random result. Validity and reliability jointly define the overall quality of assessment and are considered to be key indicators of the quality of a measuring instrument.

Both validity and reliability are important for any measuring tool, and they should be always empirically tested together (i.e. the assessment tool must not only appear valid but be reliable as well; (Crossley et al., 2002a). Likewise, Snell and colleagues (2000) emphasised the necessity of obtaining reliable, valid, and feasible assessments. Reliability is considered to be the upper limit of validity in an assessment process (Streiner and Norman, 2008). An examiner requires reliable scores; if test scores are unreliable, the test will be of very little benefit and cannot be used in making accurate pass/fail decisions (Norcini, 2000). However, an instrument can be reliable without being valid: for example, a scale that has been calibrated incorrectly will reliably report the same incorrect weight values. Likewise, testing counselling skills of pharmacy students using a multiple-choice test may result in reliable scores, but it may test their pharmacological knowledge rather than their communication skills, and thus not be a valid test of counselling. (Kimberlin and Winterstein, 2008).

Recent views of clinical teaching assessment showed that validity and reliability are frequently misunderstood and misapplied (Beckman et al., 2004). The primary focus of the examiners should be on validity of the instrument used. However, when the reliability of an assessment cannot be accurately measured because of the type of the assessment or small sample size, we should look for other evidence of the utility of

that assessment, such as its educational impact, acceptability, to determine its validity.

This approach is used by the GMC to enhance the quality of medical institutes' assessment programmes (GMC, 2010b).

3 Systematic review of empathy measures for health professionals

3.1 Introduction

This chapter describes a systematic review of the literature on empathy measures. It outlines the search strategy with a list of the selection criteria for publications and the search procedure. This is followed by a description of how data were extracted from included articles, which in turn is followed by a narrative synthesis and then a description of the limitations of this review. This systematic review provided the thesis with a summary of contemporary issues concerning the usage of different instruments for graduate and undergraduate health professionals. Lastly, the findings of the review are discussed and the reason for choosing a suitable measure for this study explained.

3.2 Background and rationale

There have already been a number of recent attempts to systematically review tests of empathy conducted in different health care settings among undergraduates and postgraduates health professionals. Firstly these were considered to see if they identified a suitable measure. The review conducted by Hemmerdinger and colleagues was the first of empathy tests, in this case for student selection (Hemmerdinger et al., 2007). They performed two literature searches, one in 2005 and an updated one in 2007, for papers related to the reliability and validity of tests of empathy and emotional intelligence or emotional quotient. However, this systematic review was limited as it contained only those papers published after 1980 and in the English language. From 36 instruments described in that review, there were only eight tools that demonstrated evidence of reliability and validity: six of them were self-rated while the other two

were patient-rated and observer-rated measures. The authors of this review were interested in empathy measures with sufficient predictive validity for use as selection measures for medical school and did not fully focus on tools used in medical training and clinical care.

A systematic review published in 2008 by Yu and Kirk identified measures of empathy in nursing specifically (Yu and Kirk, 2008). The authors demonstrated many tools used in nursing research to assess empathy but discovered that most of the 20 scales reported were inconsistent and still had to be evaluated appropriately. Therefore, a year later, they conducted another review in which the rigour of measures applied in nursing search was analysed, culminating in a 'gold standard' for application in future studies (Yu and Kirk, 2009). Despite the consistent approach they used in the second review, their results were psychometrically and conceptually unsatisfactory for the 12 measures reported, and failed to identify a tool for nursing that covered all domains or reflected user perspectives. This systematic review highlighted the need for the development of user-centred empathy measures, involving users in this process and for tools to be tested with the relevant population for each of the health care settings or context (Yu and Kirk, 2009).

In contrast, Pederson (2009) conducted a critical review to assess the methodologies of the empirical research on empathy in medicine. He reported some specific limitations in the identified publications including a predominance of quantitative studies based on self-assessment approaches removed from medical practice and the patient perspective. Furthermore, he noticed the lack of a clear definition of empathy with key aspects of empathy still not assessed. He called for including the physicians' and patients' experiences and interpretations in future studies on empathy (Pedersen, 2009).

There is agreement on the need for a specific measure of empathy that is valid, reliable, and feasible to apply in clinical practice. Such a measure also should be useful in research, education and training and have the ability to reflect users' perspectives (Hojat et al., 2003, Hojat et al., 2002a, Hojat et al., 2001). The above systematic reviews were not helpful in locating such a measure. Thus, there is a need to conduct a systematic review that critically assessed the existed empathy measures and presenting details on their development and validation. Also, it is important to determine in which context each instrument was used because the psychometric properties of any scale may change accordingly (Douglas and Nijssen, 2003).

3.3 The aim of the systematic review in this thesis

The aim of the systematic review in this thesis was to find an empathy measure specifically designed for and evaluated in the clinical context that assessed empathy (as previously defined) reliably and preferably from a patient perspective. All studies describing tools used for empathy measure in health professionals were considered provided the psychometric properties of the measures were reported to inform judgment regarding the potential utility. Unlike previous reviews, studies in any health care setting or professional grouping were included. Also, the information on utility of the existing tools was investigated to help us find the most suitable measure that could be utilised for the assessment of the relational empathy for medical students specifically.

3.4 Search strategy

The present systematic review was built on the framework of the ‘Centre for Reviews and Dissemination’ (CRD) for undertaking systematic reviews. The most recent version of CRD guidance was the main source of information on conducting systematic reviews specifically in health care research (Centre for Reviews and Dissemination, 2009) and has been used widely in the UK and internationally. Accordingly, a review protocol was set out in advance to show the approach to be followed and minimise bias. The other method to improve the reporting of the systematic reviews was the checklist of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). However, the PRISMA itemised checklist was not applicable to our study as it focuses on randomised trials.

Literature searches were performed in MEDLINE, CINAHL, PubMed, PsychNet, Web of Science, and ScienceDirect databases. The search for key words ‘empathy or empathic or empathize’ were combined with the search of the words ‘measure or assessment or scale or rating or evaluation or tool or questionnaire’ to identify relevant literature published until July 2013. The search strategy details for the systematic review can be found in Appendix 1.

3.4.1 Selection criteria

Studies were selected if they met the following criteria:

- they measured empathy in the health context and assessed its association with healthcare professionals (graduates or undergraduates of medicine, dentistry, nursing, pharmacy, etc.) and/or their clients (patients or simulated patients)

- they described the development of an empathy measure and/or reported psychometric properties (e.g., test reliability and/or validity were assessed)
- they included a measure of empathy as one component
- they were published papers (not theses or dissertations)
- they were available at the local library, via electronic journals, or through inter-library loan service
- the full text publication was in English

3.4.2 Search procedure

The present study search process covered all relevant published articles to date with no search limits. The search initially identified a total of 4,228 titles and abstracts collected from all databases used after deleting duplicates. Of these, 3604 articles were excluded as not relevant, and nine were excluded as they were in non-English publications. The remaining 615 studies were then reviewed for the inclusion and exclusion criteria mentioned above and a further 552 were unmatched against the selection criteria. In addition, 4 non-English articles (two in Spanish, one in German, and one in Persian) were found to be irrelevant to this review by reviewing their English abstracts.

Another 22 studies were found by searching Google Scholar for specific empathy measures by name. This brought the total of possible studies to 81. These were then retrieved in full. Of these, 21 were subsequently excluded as they were reports and/or no tool was reported. In contrast, an additional 14 publications were included after being identified by a manual review of the references lists of studies selected. These described the original development and validation of some measures that the original

search had not identified. Finally, 74 articles were selected for detailed review; see the flow chart below (Figure 3.1).

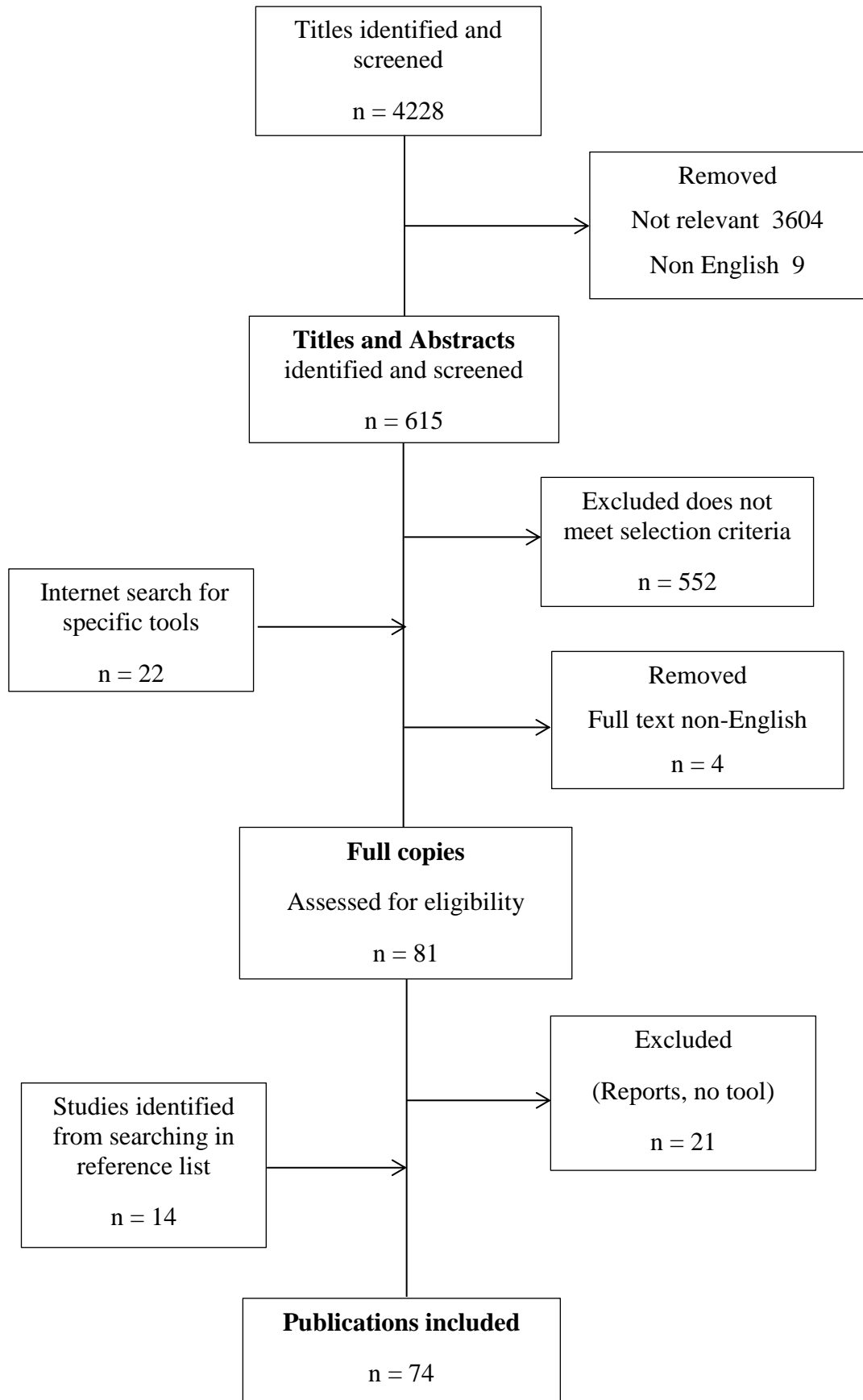


Figure 3.1: Flow diagram for the search process.

3.4.3 *Data extraction*

The process of data extraction was designed and developed to address both the review questions according to the Buscemi method of data extraction (Buscemi et al., 2006). The data extraction was performed by the principle researcher and then independently rechecked for accuracy and completeness by the principle supervisor.

The studies included in the review are listed in Table 3.1, which shows the specific empathy measure used in each study, the professional group involved and the type of assessment, and whether it was self, patient, observation, or combined. Other detailed information extracted from each paper included a summary description of each measure, the psychometric properties of the measure (validity assessments, reliability, and internal consistency), and specific comments related to that measure and its utility.

Table 3.1: Empathy assessment measures (Jefferson Scale of Physician Empathy (JSPE); Hogan’s Empathy Scale (HES); Interpersonal Reactivity Index (IRI); Empathy Communication Skills Scale (ECSS); Balanced Emotional Empathy Scale (BEES); Consultation and Relational Empathy Measure (CARE); Reynolds Empathy Scale (RES); Barrett-Lennard Relationship Inventory (BLRI); Jefferson Scale of Patient’s Perception of Physician Empathy (JSPPE); Empathy Construct Rating Scale (ECRS); Four Habits Coding Scheme (4HCS); Empathic Communication Coding System (ECCS)) and the studies included for each target population.

[illegible]

	IRI	Quince et al., 2011 Evans et al., 1993							Yarnold et al., 1996		Bellini et al., 2005		Davis, 1980; Davis, 1983a
	ECSS	Ozcan et al., 2012	Cinar et al., 2007 Ozcan et al., 2012							Ancel, 2006			
	BEES	Newton et al., 2008								Mehrabian, 1997			Mehrabian 1972; Stepien & Baernstein 2006
Patient rating	CARE								Bikker et al., 2005; Fung & Mercer 2009; Fung et al., 2009; Mercer et al., 2002a,2004 , 2005, 2007c,2008 a, 2008b, 2011; Mercer & Reilly 2004; Price et al., 2006; Wirtz et al., 2011.		Riess et al., 2011, 2012;	Mercer et al., 2008b	
Observer rating or third-party rating	4HCS								Krupat et al.,2006				
	ECCS								Bylund & Makoul, 2002				Goodchild et al., 2005
Combined self-report and client rating	RES									Reynolds, 2000; Reynolds & Brayn, 2000			

	BLRI								Barrett-Lennard, 1962				Barrett-Lennard, 1978, 1986. Gurman, 1977
	JSPPE	Kane et al., 2007											
	ECRS									La Monica et al., 1976; La Monica, 1981.			

3.4.4 *Synthesis*

Narrative synthesis was used to summarise and explain the findings of the multiple studies of this systematic review. Words and texts were used for this synthesis (i.e. adopted the textual approach to tell the story of the finding of the included studies in addition to manipulation of statistical data). The narrative synthesis proved a useful way to describe differences of tools in the included studies. However, a combination of narrative synthesis with statistical meta-analysis is often used in systematic reviews for reasons of study heterogeneity (Popay et al., 2006); (Rodgers M et al., 2009).

The papers were grouped by the type of empathy measure tool for the purposes of synthesising the findings. A narrative synthesis of the extracted data was undertaken and organised according to the different perspectives in measuring empathy (self-report, patient-rated, combined self- and patient-rated, and observer-rated).

3.4.5 *Limitations*

This systematic review has some limitations. First, like any review it is possible important literature may have been omitted. For example, not all relevant publications may be indexed in the databases used in our search and not all titles or abstract of literatures may contain the word ‘empathy’ or other related terms. This was addressed partially by the manual review of the reference lists of the preliminary selected studies. Second, the information gathered from the available relevant publications included in the study missed some validity and reliability figures of some tools as it excluded unpublished articles or dissertations. Also, this systematic review ignored the non-medically related empathy measures. Unfortunately, many publications did not present

adequate information about how empathy was measured. In many publications, the reader is referred to other sources or recommended to contact the authors.

In addition, the review only included publications in English, which may have a chance of some relevant non-English published work being omitted. However, some of these non-English papers were provided with an English abstract, although it turned out that they were not relevant to this systematic review. There were insufficient time and/or sources to track down unpublished work or to translate foreign-language papers.

3.5 Measures of empathy for health professionals

This section analyses the measures that have been developed specifically to measure empathy in the context of patient care to address the needs of healthcare professionals. The empathy measures found were listed according to the approach taken by Hemmerdinger and his colleagues (Hemmerdinger et al., 2007) in their systematic review. We found this approach helpful for the present systematic review as it helped us identify the most appropriate test for further study with medical students. The empathy measures were categorised into three perspectives: self, patient, and observer ratings (first-, second-, and third-person assessment); in addition, a group of tests of combined self-rating and client rating were considered.

3.5.1 *Self-report empathy measures*

In self-assessment or first-person measures, the assessor completes a standardised questionnaire to assess their own empathy.

3.5.1.1 Jefferson Scale of Physician Empathy

Background: Professor Hojat and colleagues established this scale at the centre for research in medical education and health care at Jefferson Medical College, University of Philadelphia. The Jefferson Scale of Physician Empathy (JSPE) was developed to measure empathy among medical students, physicians and other health care professionals (Hojat et al., 2001, Hojat et al., 2002b). It has been used in the US and translated into most of the European languages and some of the Middle Eastern languages, such as Hebrew, Persian, and Turkish, as well as into the Far Eastern languages Chinese, Japanese, and Korean.

The JSPE has been widely used across various healthcare roles including medical students (Hojat et al., 2002a, Hojat et al., 2004, Hojat et al., 2009, Hojat et al., 2011c), (Berg et al., 2011a, Berg et al., 2011b); (Chen et al., 2007, Chen et al., 2010b, Chen et al., 2012); (Kataoka et al., 2009); (Kluszcz et al., 2006); (Lim et al., 2011, Neumann et al., 2012); (Paro et al., 2012); (Roh et al., 2010); (Tavakol et al., 2011a, Tavakol et al., 2011b), nursing students (Fields et al., 2011); (Hsiao et al., 2012); (McKenna et al., 2012); (McMillan and Shannon, 2011); (Ward et al., 2012), pharmacy students (Fjortoft et al., 2011), midwife students (McKenna et al., 2011); (Kluszcz et al., 2006), dental students (Sherman and Cramer, 2005), occupational therapy students (Brown et al., 2010), and paramedic students (Williams et al., 2013) Table 3.1).

In addition, the JPSE has been used with postgraduate healthcare professionals such as physicians (Hojat et al., 2001, Hojat et al., 2002b, Hojat et al., 2003); (Fields et al., 2004); (Kataoka et al., 2012); (Kluszcz et al., 2006); (Suh et al., 2012), nurses (Fields et al., 2004); (Kluszcz et al., 2006); (Kuo et al., 2012), and residents (Hojat et al., 2001, Hojat et al., 2005a).

Description: The JSPE consists of a 20-item paper and pencil test answered on a 7-point Likert scale. Completion of the test takes between 5–10 minutes. There were two versions of this tool: one for students (the S-Version) and another slightly modified for health professionals (the HP-Version).

Psychometrics: It has been found to have satisfactory psychometric properties, with high reliability (Cronbach's alpha was 0.87 for nurses and 0.89 for physicians and medical students; (Fields et al., 2004). The construct validity was established by factor analysis, which found the major construct to be the physician's view of the world from

the patient's perspective, with a second significant factor defined as understanding the patients' experiences, feelings, and cues, both in line with the cognitive definition adopted (Hojat et al., 2001). The concurrent validity of this measure can be demonstrated as students who rated higher on clinical competence had significantly higher scores on the JPSE, regardless of speciality (Hojat, 2002a).

Utility: This instrument, which is based on self-assessment, is practical for screening large numbers of medical school applicants or in some point during medical school. However, it does not measure the actual behaviour of the practitioner or student and its validity in this regard is limited (Davis et al., 2006). Although the JSPE showed a good reliability and low validity, it stills a popular measure of empathy. Neither version of the JSPE includes any patient input, and no clear relationship between this scale and patient care (satisfaction or outcomes) has been established. There has been some progress in addressing these limitations in other recent scales, which are reviewed below.

3.5.1.2 Hogan's Empathy Scale

Background: This was the first step towards a specific measurement of empathic abilities in the general population. Hogan asked research psychologists and graduate students in psychology at the University of California, Berkeley, to describe their conceptions of a highly empathic person. The scale was based on the dictionary definition of empathy as (Hogan, 1969):

‘the intellectual or imaginative apprehension of another’s condition without actually experiencing that person’s feelings.’

It is therefore rooted in an exclusively cognitive definition of empathy. Importantly, Hogan argued that the definition of empathy involves appreciation of another's condition that does not imply (or therefore require) accuracy; consequently a self-report questionnaire was considered a suitable method of measurement.

Description: This definition was given to nine psychologists and 14 others who were asked independently to use California Q-sort items to describe a highly empathic person. A high degree of agreement in the items chosen by the psychologists and non-psychologists suggested that this definition was understandable to people outside the realm of psychology. Hogan therefore used it to develop a scale comprised of 100 items that represented characteristics most and least representative of an empathic person. Two samples of participants (100 military officers and 111 scientists/engineers) completed this new scale alongside other personality scales.

Reliability and validity analyses resulted in 64 items being retained in the final scale. The choice of sample for scale construction is interesting, as it could be argued that military officers and those in scientific careers might not be the most representative on which to base a measure of empathy for use in the general population. In addition, no information was presented in terms of gender of the samples. However, based on this analysis, Hogan developed and provided a basis for a 64-item empathy scale, known as 'Hogan's empathy scale' (Hogan, 1969). A greater empathic ability is rewarded higher scores from a raw score ranging from 0–39.

Psychometrics: The scale showed acceptable levels of internal consistency and test–retest reliability with general population and a low-to-moderate level of concurrent validity. Statistically significant differences in empathy levels between groups and

gender supported the construct validity of the scale. Of some concern, is that in an investigation of the scale with a population of healthcare professionals (occupational therapy students), internal consistency was found to be only 0.57, while test–retest reliability was only 0.41 over 12 months (Froman and Peloquin, 2001).

Utility: This scale was originally developed with the general population to assess the intellectual appreciation of the feeling of others but not specifically for healthcare professionals; in addition, it does not assess empathic behaviour (Hogan, 1969). Using a restricted definition may have limited the success of the Hogan empathy scale, as a measure as reliability has not been consistently established. Although evidence regarding validity may be viewed positively, concerns about using a purely cognitive definition remain and inclusion of an affective component might be beneficial.

3.5.1.3 The Interpersonal Reactivity Index

Background: Davis developed the Interpersonal Reactivity Index (Davis, 1983a) based on a multidimensional model of empathy which sought to include both cognitive and emotional constructions, and differentiate between personal distress and empathic concern (Davis, 1980).

Description: This self-report measure consists of four 7-item subscales (perspective taking, fantasy, empathic concern, and personal distress) and taps separate aspects of the global concept of empathy. The four distinct subscales were based on the following rationale (Davis, 1983a):

‘Empathy can best be considered as a set of constructs, related in that they all concern responsiveness to others but are also clearly discriminable from each other.’

According to Davis, while these four subscales do not exhaust the possible range of reactions to others, previous theory and research suggests they reflect the variety of reactions that have at some point been referred to as empathy. To establish validity of the Interpersonal Reactivity Index and four subscales, Davis investigated relationships between these and other psychological constructs. Findings from construct and convergent validity studies provide support for the theoretical and psychometric properties of the scale (Davis, 1983a).

Psychometrics: Although the Interpersonal Reactivity Index has proved a popular instrument based on a multidimensional theory that includes both emotional and cognitive components, it was not designed specifically for the healthcare context. However, Yarnold et al. (1996), in a study of 114 physicians and 95 medical students, found the Interpersonal Reactivity Index to have good structural integrity and convergent validity, with a measure of problem solving in the context of concern for the well-being of others. Evans et al. (1993) also undertook a study using the Interpersonal Reactivity Index to assess empathy in 55 medical students during clinical training. After completing the measure, students’ behaviours in a 20-minute consultation were scored, using five items from the 16-item History-taking Interview Rating Scale that assess behaviours relevant to empathy such as eye contact, use of jargon, etc. Scores on the Interpersonal Reactivity Index and the five items measuring empathic behaviours were positively correlated. The measure has been used

specifically with healthcare professionals such as medical students (Quince et al., 2011) and internal medicine residents (Bellini and Shea, 2005).

Utility: The Interpersonal Reactivity Index, though not developed for health care professionals, offers a multidimensional measure clearly including both cognitive and emotive aspects of empathy and appears to correlate with relevant behavioural assessment of consultation skills. However it remains a self-completed measure with the associated limitations for assessment purposes.

3.5.1.4 The Empathy Communication Skills Scale

Background: This instrument was developed by Dökmen in late 1980s in Turkey, to measure the affective and cognitive components of empathy along with the verbal response dimension of communication (Ancel, 2006). There is also a version called the Empathic Skill Scale B-form, which was used to evaluate nursing students (Cinar et al., 2007) and to determine the effect of educational training programmes to improve empathic skills for medical and nursing students (Ozcan et al., 2012).

Description: In the scale, six health problem scenarios were presented along with 12 potential response-sentences listed under each problem. The participants were asked to choose four responses which they liked best (one of the 12 responses was an irrelevant response included in order to determine the random replies). The students chose 24 responses, four for each of the six problems; and the point for each response they chose was given according to the evaluation section of the scale. The maximum point collected through the Empathy Communication Skills Scale was found to be 219, while the minimum was 64. The higher grade is the higher level of empathy.

Psychometrics: The validity and reliability studies for the Empathy Communication Skills Scale were carried out by Dökmen. The scale was applied to a group of 60 university students and 24 psychologists from different institutions. The instrument was applied on 60 subjects with a two-week interval to test the reliability which found to be ($r = 0.83$, $P < 0.001$). For the validity study, significant differences were found between the two groups ($t = 8.15$, $P < 0.001$), with favour to the psychologists.

Utility: This measure was developed for nursing and for the development of empathy in nursing students by using psychodrama. The Empathy Communication Skills Scale appears to offer potential as a paper-based assessment of empathic responses that has relevance to health care practitioners and includes emotive and cognitive elements. It may be less vulnerable to ‘gaming’ than some of the previously reported scales due to the different response format. However, it remains distanced from patient interactions and may not reflect how students or clinicians respond in practice.

3.5.1.5 The Balanced Emotional Empathy Scale

Background: The Balanced Emotional Empathy Scale represents an updated version of the Questionnaire Measure of Emotional Empathy (QMEE) which was developed by Mehrabian and Epstein (Mehrabian and Epstein, 1972). This instrument was used later on by Newton to monitor the levels of vicarious emotional empathy (i.e. to have a visceral empathic response, versus role-playing empathy) in medical students (Newton et al., 2008).

Description: The Balanced Emotional Empathy Scale includes 30 items (positively and negatively worded, 15 in each category) in which the participants are asked about their agreement or disagreement with a 9-point scale (Mehrabian, 1997a).

Psychometrics: The Balanced Emotional Empathy Scale measure has good internal consistency and test–retest reliability (Cronbach’s $\alpha = 0.87$, $r = 0.77$). The measure showed a high positive correlation of 0.77 with the Emotional Empathic Tendency Scale (EETS), and the correlation coefficient was 0.67 (Mehrabian, 1997b).

Utility: The Balanced Emotional Empathy Scale measure is easy to administer and score. Recently, administering, scoring, and interpreting this measure can be done through Windows software available from the author. However, the Balanced Emotional Empathy Scale was not originally developed in a healthcare setting and was designed to assess purely the emotive component of empathy (Stepien and Baernstein, 2006). This scale has robust psychometric properties and may offer potential when combined with a cognitively oriented empathy assessment. Like other self-report tools it may not relate to consultation behaviours, and this has not been tested to date.

3.5.2 Patient-rating empathy measures

In patient rating, also known as second-person assessment, patients assess their carers’ empathy after a clinical interaction.

3.5.2.1 The Consultation and Relational Empathy Measure

Background: The Consultation and Relational Empathy (CARE) measure was developed on the basis of arguments that patients’ views are central to the effectiveness of empathy in the clinical encounter (Mercer et al., 2004). The authors also intended to build on the work of Reynolds (see below) by creating a measure of empathy which could be relevant for clinical encounters other than nursing (Mercer and Reynolds, 2002a). Their initial measure was piloted using a sample of general practitioners and patients, and qualitative and quantitative examinations of validity

allowed for revision of the CARE measure until the third version of the scale was deemed satisfactory.

Description: The CARE measure is a ten statement questionnaire which is answered by patients using a five-point scale from poor to excellent (plus a ‘Does not apply’ option) immediately after the consultation. This measure focuses on the behavioural dimension (it was intended for use by patients appraising medical professionals) rather than a multidimensional model of empathy.

Psychometrics: The CARE measure was validated by Mercer and his colleagues at the University of Glasgow and the University of Edinburgh (Mercer and Reilly, 2004, Mercer et al., 2004, Bikker et al., 2005, Mercer et al., 2005b). The final version of the questionnaire was completed by 3,044 patients for 26 GPs from different practices. Most of the patients rated the items within the measure as being very important to their consultation with the doctor. This measure showed a high internal reliability (Cronbach’s $\alpha = 0.92$) and was meaningful to patients from different social classes (Mercer et al., 2004). The acceptable reliability of the CARE measure was achieved with 50 patients per doctor. The final version of the CARE questionnaire was correlated with the Reynolds Empathy ($r = 0.85$) and Barret–Lennard Empathy ($r = 0.84$) scales and showed high face and content validity (Mercer et al., 2005b). These findings supported the feasibility and reliability of the CARE measure in primary care settings in the UK. In addition, this measure was revalidated and used after translation to other different languages such as Chinese (Fung and Mercer, 2009a, Fung et al., 2009); (Mercer et al., 2011b); (Wirtz et al., 2011).

Utility: The CARE measure is useful for assessing empathy in general practice consultations. It was used successfully beside other tools assess patient satisfaction and

evaluate the effect of increasing consultation length on patient enablement in general practice (Mercer et al., 2007c). In secondary care, this measure was used to evaluate acupuncture patients' perceptions of practitioner empathy (Price et al., 2006) and to measure empathy level changes of otolaryngology residents (Riess et al., 2011, Riess et al., 2012).

In addition, Mercer and Douglas have piloted the CARE instrument among ten specialities in a single hospital trust in Scotland. The CARE measure proved its usefulness and high relevance within 1,105 out-patients attending 25 consultants. Recently, the CARE measure was included in Murphy's study on workplace-based assessment tools and can be used in high-stakes assessment and in secondary care (Mercer and Murphy, 2008). To date, this measure has not been evaluated in undergraduates or other professions.

3.5.3 Observer-rating empathy measures

In observer-rating empathy measures, an observer (third person) uses standardised assessments of empathy to rate interactions between healthcare providers and patients or simulated patients (standardised patients).

3.5.3.1 The Four Habits Coding Scheme

Background: This instrument was based on the Four Habits Model which was used for teaching and improving clinicians' communication skills in southern California.

Description: The Four Habits Coding Scheme identified a set of 23 items/behaviours, each associated with one of the 'Four Habits.' The first habit (Invest in the Beginning) contains six items that focus on creating rapport quickly and planning the visit (e.g.

demonstrating familiarity with the patient and greeting the patient warmly). The second habit (Elicit the Patient's Perspective) contains three items (eliciting the patient's understanding of the problem, understanding the patient's goals for the visit, determining the impact of the problem on the patient's life). The third habit (Demonstrate Empathy) contains four items that deal with encouraging, accepting, and responding to the patient's expression of emotion (e.g., helping patients to identify their emotions; using appropriate non-verbal behaviour). The fourth habit (Invest in the End) contains ten items that focus on effective decision making and information sharing (e.g. testing for comprehension and determining the acceptability of the treatment plan; (Krupat et al., 2006).

A scale of five points rated each the 23 items of this measure. The users of this measure provide each rater (coder) with 8–10 hours of training in order to reach an acceptable level of reliability. Coding is done on video tape plus 2–5 minutes of scoring time for each interaction.

Psychometrics: The Cronbach's alphas for the four habits (1–4 respectively) were 0.71, 0.51, 0.81, and 0.61. The overall inter-rater reliability was a generally acceptable 0.72. In terms of construct validity, several correlations were found among components of this tool and the Roter Interaction Analysis System (RIAS) categories, such as ratings of non-verbal behaviour, back channel response, and visit time (Krupat et al., 2006).

Utility: The Four Habits Coding Scheme was developed from a consultation model without explicit patient contribution and is not explicitly anchored to empathy per se. It required considerable assessor training and was time consuming.

3.5.3.2 Empathic Communication Coding System

Background: Bylund developed this instrument and tested its validity and reliability in 2001. This instrument does not intend to measure empathy as an internal process but to measure observable behaviour only. Its ability is limited to physician verbal behaviour on empathic communication levels and does not include non-verbal behaviours. Also, it is important here to remember that empathy differs in some ways from empathic communication, as a doctor may feel empathy but still not be ready or able to show it during communication with patients. The Empathic Communication Coding System measures the actual communication behaviours of physicians but not their intentions and interpretations resulted from communication with patients.

Description: The measure has two parts: it starts with an identification of empathic opportunity and then physician responses within videotaped interactions are coded. The purpose of part one is to provide a confined list of patient statements revealing windows of opportunity (instances) during which patients discuss their personal and/or emotional concerns. The second part is a system to code physician responses to empathic opportunities. This measure required selected videotaped encounters from physicians until adequate empathic opportunities are presented for appraisal.

Bylund conducted two questionnaire-based studies: the first was to learn about physician behaviours that patients perceived to be empathic (creating a hierarchy of empathic communication behaviour), and the second was a validation test to attend to these issues (Bylund and Makoul, 2002). The more empathic the professionals' response to emotional opportunities, the more satisfied patients were with their consultations ($r = 0.41$, $d.f. = 15$, $P = 0.05$; (Goodchild et al., 2005).

Psychometrics: Unfortunately, the details on the reliability and validity of this instrument were only reported in Bylund's unpublished doctoral dissertation and thus could not be obtained.

Utility: The Empathic Communication Coding System may offer a reliable and conceptually derived observational measure of physicians' or students' empathic behavioural responses. However, it requires further validation and is a highly labour intensive approach.

3.5.4 Combined empathy measures

Combined measures comprised those that showed the ability for multiple usage, as self-rated, client-rated, or observer-rated, under different circumstances.

3.5.4.1 The Reynolds Empathy Scale

Background: This scale was developed in the UK as part of a nursing research programme and measures empathy in clinical nursing training. In order to develop the scale, 30 patients were asked for their perceptions of effective and ineffective interpersonal behaviours demonstrating empathy as identified from the academic literature (Reynolds, 2000, Reynolds and Brian, 2000). To be included in the scale, the item must have been mentioned by at least 20 of the 30 patients.

Description: The outcome was a 12-item questionnaire (six positive and six negative) through which the attitude or behaviour of a counsellor (e.g. nurse) is evaluated. The rater asked to give an opinion on every statement and decide the degree to which the rater perceives the person that he/she rating (e.g. the rater himself, a nurse, or an associate, etc.) according to a seven-point scale from 1 (Always like) to 7 (Never-like).

In addition, the rater was guided to read each statement on the empathy instrument and consult the operational definitions and clinical examples (provided in the User's Guidelines) before scoring the instrument and to tick one response for each item on the scale. The Reynolds Empathy Scale was thus developed as an interview scale designed to measure the patient's perception of nurse's empathy through observing the verbal and non-verbal interaction between nurses and patients (Reynolds, 2000).

Psychometrics: The internal consistency of the Reynolds Empathy Scale was high (Cronbach's $\alpha = 0.90$) and the test-retest reliability was also high (0.90) after over four weeks. This measure was correlated with La Monica's Empathy Construct Rating Scale, and no factorial analysis was reported (Reynolds, 2000). The face and content validity of the Reynolds Empathy Scale were confirmed by six experts from nursing and clinical psychology.

Utility: The Reynolds Empathy Scale is a useful assessment tool for clinical and research settings. It is used in nursing teaching in several countries and offers enhanced realism as it is based upon face-to-face clinical interactions. However, it failed to incorporate a multidimensional model of empathic processes. Patients were not involved in assessment of their perception of nurses' empathy, although their views were considered in generating the scale items.

3.5.4.2 The Barrett-Lennard Relationship Inventory

Background: This measure was designed to investigate changes in the clinician-client relationship in the psychotherapeutic context. The specific items of this measure were derived from Roger's Conditions of Therapy and Bown's existing Relationship Sort (Barrett-Lennard, 1962). The groups of items representing each variable were revised by five judges from the University of Chicago Counselling centre. They eliminated 92

items from the original inventory to obtain 64 items divided into four subjective subsets of interpersonal relationship ('Empathic Understanding', 'Level of Regard', 'Unconditionality', and 'Congruence'; (Barrett-Lennard, 1986).

Description: A multiple-choice questionnaire form with three grades of 'Yes' and 'No' responses to 64 items. The instrument can be completed by either the client using the client form (Other-to- Self) or it can be completed by the therapist using therapist form (Myself-to-the-other), which is equivalent to the client form and worded in the first person for the therapist to describe their response to their clients (Barrett-Lennard, 1986).

Psychometrics: The internal reliability of the Barrett-Lennard Relationship Inventory was reported by Gurman, who collected data from 14 studies (four of them used Cronbach's alpha coefficients, the others used split-half reliabilities) on this instrument and found that the mean internal reliability of the Barrett-Lennard Relationship Inventory in these studies was 0.91 (Gurman, 1977), with subscale inter-correlations mean $r = 0.45$, clients' mean $r = 0.65$, therapists' split-half reliability $r = 0.86$, clients' $r = 0.96$, and therapists' test-retest reliability $r = 0.89$ over four weeks. The difference between expert and non-expert therapist was statistically significant (no P -value reported) and the agreement between these two pairs of therapists was significant ($P < 0.01$; Gurman, 1977). The content validity of the Barrett-Lennard Relationship Inventory was supported by a procedure conducted by five experts to eliminate the non-differential items (i.e. items that did not express the variable they were designed to represent; (Barrett-Lennard, 1978).

Utility: The Barrett-Lennard Relationship Inventory was used in actual and analogue therapy settings as well as to evaluate other types of relationships such as those with

teachers, parents, and friends. In addition, it has been used widely in psychotherapy research (Barrett-Lennard, 1978). As this measure has a long list of items, it tended to be used by various researchers as isolated subscales and in different forms with modifications of content and response format rather than the whole inventory. This has posed significant challenges to its further empirical validation.

3.5.4.3 Jefferson Scale of Patient's Perception of Physician Empathy

Background: This measure was developed by Hojat to investigate the relationship between physicians' self-report scores on the JSPE and their patients' perceptions (Hojat, 2001). It is intended to measure patient's perceptions of his/her physician's empathic concern and understanding.

Description: Patients answer five items on their interaction with their physician using a 5-point Likert-type scale (1= strongly disagree to 5= strongly agree) confidentially in the waiting room after completing their consultation.

Psychometrics: The reliability coefficient of Jefferson Scale of Patient's Perception of Physician Empathy was relatively low. The patients' scores were highly positively skewed, and the correlation coefficient with physicians' self-ratings on the JSPE was non-significant 0.24 (Kane et al., 2007).

Utility: The psychometric properties of this scale suggest that it is not currently fit for further evaluation without additional development.

3.5.4.4 Empathy Construct Rating Scale

Background: The La Monica's Empathy Construct Rating Scale is an empathy measure that is specific for healthcare. It was originally developed to evaluate an

empathy training programme for nurses in hospital settings (La Monica et al., 1976) (La Monica, 1981). The first version of this scale included 259 items which were generated by 25 female graduate nursing students. These items were reduced to the final 100 via an item facility analysis using three experts (psychometrics, psychology, and nursing) and ten students' ratings.

Description: The Empathy Construct Rating Scale is a self-scored instrument that can be used to measure empathy of the person himself/herself or for another using a six-point scale. Participants are asked to grade their responses from 'extremely unlike' to 'extremely like' after reading each statement.

Psychometrics: No reliability statistics for the self-rating version were reported. This tool exhibited high level of internal consistency and split-half reliability. The discriminant validity was weak and the convergent validity was not evident. Also, there was insufficient evidence on face and content validity of this scale.(Reynolds, 2000).

Utility: The Empathy Construct Rating Scale is a comprehensive tool but lengthy and developed without patient input. It also lacks sound construct, convergent or criterion-related validity. It was developed for nurses only and has been employed in nursing research with lack of nurse–patient interactions (Yu and Kirk, 2009).

3.6 Results of the systematic review

Empathy was measured from three different perspectives:

- 1) the self-rating assessment of empathy using standardised questionnaires completed by those being assessed;
- 2) the patient rating (or second-person assessment), with the use of questionnaires given to patients to assess the empathy they experience from their carers; and
- 3) the observer rating, with the use of standardised assessments by an observer to rate empathy in interactions between health personnel and patients, including the use of 'standardised' or simulated patient encounters to control for differences between patients or case mix.

This systematic review identified 12 different tools: five were self-report assessments, one was patient rated, two were observer-rated assessments, and four were a combination of self-rated and client-rated. The target populations for these tools were distributed between undergraduate and postgraduate health professionals from different disciplines (doctors and nurses primarily).

The psychometric evidence for the reliability and validity for each empathy measure was collated as far as possible from all papers available using a data extraction forms. The results of this review revealed the domination of self-report measures of empathy within the health context such as the JSPE, Hogan's Empathy Scale, the Interpersonal Reactivity Index, the Empathy Communication Skills Scale, and the Balanced Emotional Empathy Scale; the others concerned empathy from the patient perspective (CARE) or involved an observer assessment of empathy (e.g. the Four Habits Coding Scheme, the Empathic Communication Coding System), four measures included both

self and client ratings (the Reynolds Empathy Scale, the Barrett-Lennard Relationship Inventory, Jefferson Scale of Patient's Perception of Physician Empathy, and the Empathy Construct Rating Scale). The JSPE and CARE scales were the most well-evaluated measures and have been used in many studies and in different settings. Some other tools have only been used in few studies (e.g. the Barrett-Lennard Relationship Inventory, the Reynolds Empathy Scale, Hogan's Empathy Scale, and the Four Habits Coding Scheme).

The validity of most measures was addressed to some extent, while other tools showed more than one type of validity (e.g. CARE, the Empathy Construct Rating Scale, Hogan's Empathy Scale, JSPE, and the Reynolds Empathy Scale). Face and content validity were evaluated by a panel of experts (e.g., the Barrett-Lennard Relationship Inventory, the Empathy Construct Rating Scale, Hogan's Empathy Scale, the Interpersonal Reactivity Index, JSPE, and the Reynolds Empathy Scale) and patients (CARE). Construct validity was the most frequently reported method and was confirmed by convergent validity and discriminant validity.

Validity assessments of self-reported measures were primarily concerned with assessing the relationship between measured empathy and various aspects of the consultation or clinical knowledge. None of the first-person measures were validated by directly comparing assessments with empathy as judged by patients, although JSPE was subjected to a test of predictive validity through correlating empathy scores with later ratings of empathy from directors during residencies.

Concurrent validity was sought through comparison with other empathy measure such as Hogan's Empathy Scale and the Reynolds Empathy Scale or by predictive validity through assessing the ability to predict future changes (e.g. the Barrett-Lennard

Relationship Inventory). Reported criterion validity was low, with the exception of the Reynolds Empathy Scale.

Reliability data were presented for most of measures, and internal consistency was the most frequently used method. The patient-rated measure (CARE) showed excellent internal consistency. In addition, self-rated measures generally had adequate internal consistency. Most tools had a moderate to high level of internal reliability, with Cronbach's alphas ranging from 0.70 to 0.98 and split-half correlation coefficients of more than 0.84, though subscale correlations were low or not statistically significant (Interpersonal Reactivity Index). Moderate to high reliability was shown, with correlation coefficients ranging from 0.61 (Interpersonal Reactivity Index) to 0.98 (Empathy Construct Rating Scale). Equivalence was reported for the Reynolds Empathy Scale, as demonstrated by inter-rater reliability. The initial agreement between raters for the Reynolds Empathy Scale was low, but the final agreement reached 41.6–91.6%. The Empathy Construct Rating Scale involved third-party ratings, but evidence for inter-rater reliability was absent.

3.7 Choosing a suitable measure for this study

Selection of an assessment measure should be first made within limitations of the available resources such as time, money, personnel, and expertise. Having reviewed the main available empathy assessment tools, I had to choose the most suitable measure for this study. The systematic review showed that most measures were based solely on self-reports. However, self-assessment of empathy is particularly difficult and may result in overlooking the students that are at most need of training or supervision (Davis, 2006). Hence, though reliable and cost effective, these were discounted as unfit for the purposes required.

The suitability of a particular type of any measure depends to a large extent on the situation in which it is to be used. For example, observer-rated measures are considered unpractical for screening large number of medical school applicants and require the rater to make (more or less valid) assumptions about the patient's feelings or experiences. Hence measures such as the Four Habits Coding Scheme and the Empathic Communication Coding System need extensive training, video facilities, etc. So, although these tools can offer reliable and valid measures of behavioural aspects of empathy, they are expensive and challenging to implement.

In Section 2.4.3, the importance of patients' perceptions was shown because their rating of clinicians' empathy correlated more highly with clinical outcomes. Indeed, it may be argued that the patient's rating might be the only type of assessment that can evaluate patients' appreciation of the health professionals' empathic behaviour (Barrett-Lennard, 1981). Unfortunately, the systematic review revealed that patients have been infrequently asked to evaluate medical students' or doctors' empathy and that simulated or standardised patients are more commonly used.

The last decade or so has shown significant progress in the development of empathy scales for use in healthcare contexts. For example, the publication of the Reynolds Empathy Scale (Reynolds, 2000) and the CARE measure (Mercer et al, 2004) have marked significant developments in efforts to assess empathy based on patient perceptions of healthcare professionals. The CARE measure is intended to be of use to a range of healthcare professionals, while the Reynolds Empathy Scale is nursing specific.

Medical schools' assessments systems lack opinions from real patients on future doctors (senior medical students), and the CARE measure is the only instrument that provides a mechanism for patients to rate the students' relational empathy during independent clinical consultations. Thus, the CARE questionnaire offers an independent perspective which could be used as a complement to existing instruments for the assessment of communication skills in the whole programme (Schuwirth and Van der Vleuten, 2004b). It covers an important field of person-centred approach, meets the emerging requirement to include patients in the process of evaluating professionals, appears reliable, and is feasible for use in testing among students.

For all the reasons mentioned above and the information gathered via this systematic review, a decision was reached to choose the CARE instrument to measure empathy for senior medical students and to evaluate its validity and reliability within undergraduate medical education.

4 Methods used for the utilisation of the CARE measure

This chapter describes the methods chosen for testing the validity and reliability of the CARE measure for use with medical students. It outlines the research design, data collection, analytic approach, and ethical considerations. The CARE questionnaire and changes made to the original form are described and the reason for their adoption outlined. This chapter presents student and patient recruitment and selection methods as well as the process of questionnaire distribution and the role of practice managers, tutors, and students in collecting the data.

4.1 The study design

The reliability and validity of the CARE measure among graduate doctors was previously demonstrated (the details and the facts on this measure will be presented later in this chapter). The next step therefore, is to assess the reliability and validity of the CARE measure with undergraduate medical students (Fung and Mercer, 2009b), in this instance, during their GP attachments. The intention is to specifically assess if the CARE measure can offer a reliable, valid, and practical means of providing patient feedback on the consultation and humanistic skills of medical students in the clinical phase of their training.

The CARE questionnaire was originally designed to ask actual patients (as opposed to simulated) to rate performance on several dimensions related to empathy immediately following a clinical interaction. This study employs the same approach with senior medical students on general practice attachments assessing consultations with real patients in which students had the opportunity to lead the interaction. The idea was that the CARE measure, if a reliable, could benefit their training by improving the

assessment of key aspects of being a good doctor. It would help to identify the need for additional support and to tailor the support to students who demonstrate low empathy skills in particular.

This aspect of the study sought to address the following questions:

- Is the CARE questionnaire feasible in undergraduate GP attachments?
- What levels of reliability can be achieved with the CARE measure in this context?
- What is the concurrent validity of the CARE instrument in undergraduate medical students?
- What practical factors arise from the use of the CARE questionnaire as a patient feedback measuring tool?

4.2 Ethics Consideration and approval

This study was focused on students but also included NHS patients, so potential ethical concerns could have required separate approvals. Consultation with the local NHS Research Ethics committee indicated they considered this a matter for the university to consider. Therefore, the view of the Dundee's University Research Ethics Committee was sought by formal application. However, they concluded that it was an evaluation of normal teaching practice, without any patient identifiers, and did not require formal approval (Appendix 2).

Professor Mercer was contacted for his agreement prior to the use and modification of the CARE questionnaire for this purpose. In addition, a number of steps were taken to ensure that no ethical concerns were likely to emerge. These are outlined below.

4.2.1 Patients' ethical issues

Patients may have had concern that their ratings could be seen by the student or other staff. Therefore, it was emphasised in the instructions that forms should only be completed after the consultation and outside the consulting room, and that completed questionnaires were to be returned sealed in the envelopes provided. Additionally, the instructions to the practice managers reinforced the ethical necessity of maintaining these safeguards. Agreement of the patients to complete the questionnaire was considered to represent consent.

4.2.2 Students' ethical issues

Some students could have been worried about their scores, especially those from the first year of the study. The concern was being that poor scores could have an effect on students' academic rating. A second ethical consideration for students was that they might experience negative thoughts or emotions as a result of viewing their personalised results.

These concerns were minimised by reinforcing adherence to the instructions and caveats included in the instruction form to emphasise that the validity of CARE measure for use with medical students had not been proven yet. Thus, students' scores were only provided to the students themselves, and they were free to decide whether or not to share these with their tutor. After the first year, students were presented with reference ranges to help them assess their performance in the context of their peers. Tutor training was offered at staff development, and students were invited to discuss concerns with their practice tutor or the attachment lead.

4.3 Background on the CARE measure

A clear view on background and use of this tool is important for understanding the reasons behind utilising the CARE measure in this study. So this section outlines the background and intentions behind the use of this instrument. We will demonstrate how the CARE measure was developed, and where and when it was first used. In addition, this section will describe the steps of preliminary validation and reliability testing this measure underwent in addition to exploration of the relevance and use of the CARE measure in general practice. This section will include research in the UK and globally concerning the CARE measure in different clinical settings and with medical students.

4.3.1 Development of the CARE measure

Although several measures to assess empathy in psychiatric or in nursing settings (in secondary care; (Burns and Auerbach, 1996); (Horvath and Symonds, 1991) had been developed, these failed to reflect the clients' views or were more suited to less differentiated practice, as experienced by medical students in their general training (Reynolds, 2000). As described and discussed in Chapter 3, the CARE measure offered most promise in this context, with the added advantage that patient views on students are being actively encouraged by the GMC at present.

The CARE measure was developed in 2002 by Mercer and his colleagues from the Departments of General Practice in the Medical Schools of Edinburgh University and Glasgow University (Mercer and Reynolds, 2002a). The CARE measure showed high internal reliability and was meaningful to patients from different social classes. There was high correlation between Mercer's new measure with Reynolds Empathy Scale and the Barret-Lennard Empathy Scale as they were focusing on patients' perception and behavioural dimension of empathy. In addition, it showed a significant correlation

with the Patient Enablement Index (Mercer and Reynolds, 2002a). So, initial work suggested it was a worthwhile instrument.

4.3.2 Preliminary validation of the CARE measure

The initial version of the CARE measure was preliminary validated by Mercer and his colleagues in the 2004, based upon a broad definition of empathy in the context of clinical encounter. The researchers reviewed the commonly used measures of empathy and their work was supported by Mercer's previous qualitative research on patient's views on holistic care (Mercer et al., 2004a). In addition, Mercer et al. sought advice and feedback from experts on consultation measures and on empathy in nursing, to support the face and content validity of CARE (Mercer and Reilly, 2004). The first pilot study showed high correlation with the Reynolds Empathy Scale ($r = 0.85$) and Barret-Lennard Empathy Scale ($r = 0.63$). Based upon feedback from patients, GPs, and their own analysis, Mercer and colleagues modified CARE by reducing number of items from 13 to 10 and changing the rating scale to five points instead of ten points to reduce the skewed distribution. Other changes concerned the wording of the questions were made, and a short explanatory description was added (Mercer et al., 2004a).

A second study was conducted confirming the same strong correlation with the Reynolds Empathy Scale and Barret-Lennard Empathy Scale measures but with a less skewed distribution at this time ($\text{skew} = -0.634$). The high level of internal reliability of this version (Cronbach's $\alpha = 0.92$) and patients' feedback on the revised version indicated no further major modification was required. Therefore, a third study was carried out in which the final version of the CARE measure (Appendix 3) was tested against the Reynolds Empathy Scale ($r = 0.85$) and the Barret-Lennard Empathy Scale

($r = 0.84$) which confirmed its concurrent validity, along with ten patient interviews to support the face validity (Mercer et al., 2004a).

Later in 2004 after the development and preliminary validation of CARE, Mercer and his colleagues described the relevance and practical use of the measure in routine general practice and demonstrated the ability of the CARE measure for discriminating between doctors in terms of patient interactions. Research has now been conducted into performance on the CARE measure based on over 3,000 GP consultations in areas of high and low deprivation in Glasgow (Mercer et al., 2005b). The majority of the patients endorsed the importance and relevance of the CARE measure for their consultation in general practice. After completing the ten items of the CARE measure, patients were asked to give their view on doctors' attitude and skills listed in the questionnaire if they were important to them. At the same time, the GPs were directed to rate the importance of these items from their perspective. This study was strengthened by the wide distribution of the patient sample, high response rate, and low number of patients reporting that the items of the measure were not applicable to their case. This also supports the face validity of the CARE measure. In addition, Mercer and colleagues used G-theory to demonstrate that a minimum of 50 patient-completed questionnaires were needed per doctor to obtain a reliable estimate for their mean CARE measure score (>0.8), the level generally accepted for high-stakes assessments.

In secondary care, Mercer and Douglas piloted the CARE instrument among ten specialities in a single hospital trust in Scotland. The CARE measure proved its usefulness and relevance within 1,105 out-patients attending 25 consultants. The CARE measure items were considered of major importance for around 90% of patients. The internal reliability was high (Cronbach's $\alpha = 0.94$), and an overall

reliability coefficient was achieved with 40 patients per doctor. In addition, the CARE measure showed high face validity and there was little influence of socio-economic factors on this measure. Their findings supported the feasibility and reliability of the CARE measure in secondary care as well as primary care (Mercer and Murphy, 2008).

The CARE measure was used successfully beside other tools to determine the doctor satisfaction with time and overall satisfaction, as well as to evaluate the effect of increasing consultation length on patient enablement in general practice, particularly those with complex needs in an area of extreme deprivation in Scotland (Mercer et al., 2007c). In addition, recently, the CARE measure was included in Murphy's study on workplace-based assessment tools. The only change he made was to use a seven-point rating scale (Murphy, 2009a). The reliability of the CARE measure was very similar to that found previously by Mercer (2005).

4.3.3 Current use of the CARE measure

The CARE measure is considered as the empathy measure of choice for healthcare staff in NHSScotland and forms part of the quality strategy of the Scottish government to obtain feedback from patients and appraise physicians (NHSScotland, 2010). It has utility not only in research but also as a tool for self-audit because it provides doctors with direct feedback on their strengths and weaknesses; in terms of relational empathy, as perceived by their patients. Therefore, since 2003, NHS Education Scotland, with the Royal College of General Practitioners, has adopted CARE measure for use in GP appraisal and revalidation in Scotland (RCGP Scotland, 2003) and have offered a web-based feedback system since 2006 (RCGP Scotland, 2006). In addition, it is used in three significant training contexts:

- GPs Modernising Medical Careers (MMC) 2007, Development and maintaining an assessment system (GMC, 2007);
- GPs seeking Royal College membership (iMAP) (RCGP, 2011);
- GP registrars (Murphy et al., 2009b) as a compulsory component and part of work place-based assessment since 2007.

In addition, the CARE measure is being used with the Consultation Quality Index (CQI_2) in partnership between the (Department of Health, Diabetes UK, NHS Diabetes) and the Health Foundation, and has been piloted with anaesthetists (Mercer, 2009).

Nowadays, many the UK healthcare institutions are utilising the CARE measure in different ways (Mercer, 2009) For example, in Glasgow University, the CARE measure is used for the assessment of nurses' communication in 'Keep Well', the Scottish government's flagship health inequalities intervention, and in the University of Newcastle, the measure is used in a relational approach to decision-making support in consultation.

4.3.4 International use of the CARE measure

The CARE measure has been translated into many languages other than English and used with GPs from different nationalities. New versions of the CARE questionnaire has been created and used around the world. Here we mention some examples of these versions.

The Chinese version of the CARE measure was created by Fung and Mercer in Hong Kong. The authors utilised a translated version of the CARE measure to explore the Chinese patients' view on quality of primary care consultations. The Chinese patients

showed interest in the translated CARE questionnaire, which encouraged researchers to utilise it in a Chinese population. They found also that the CARE measure items were matched well with the themes relating to consultation process in Hong Kong (Fung and Mercer, 2009b).

The results of that study encouraged the researchers to conduct further quantitative validation. Accordingly, a second study was conducted by Fung and colleagues assessed the reliability and validity of this Chinese version of CARE and supported its validity and reliability within primary care in Hong Kong. They found that the Chinese-CARE measure was highly acceptable to patients, had high face validity, and supported construct validity by significant hypothesized relationships with other variables such as patient enablement and patient satisfaction. In addition, the study results showed very high internal reliability (Cronbach's $\alpha = 0.96$; (Fung et al., 2009).

Moreover, the Chinese version of the CARE measure has proved its reliability, acceptability, and feasibility in differentiation between the interpersonal competencies of the Chinese doctors in primary care settings in Hong Kong (Mercer et al., 2011a). In this recent study, the Chinese version of the CARE measure showed very high internal consistency (coefficient = 0.95) and only 15–20 patients were needed to differentiate between doctors (inter-rater reliability >0.8). This version showed an association of general health with the nature of the health problem (e.g. acute or chronic) which had not been found in the original version of the CARE measure.

Similarly, Neumann and colleagues at the Universities of Cologne and Friburg translated the CARE measure to German and validated that German-version in study on inpatient cancer patients (Neumann et al., 2008). In addition, the CARE measure

has been included in many assessment systems in different countries other than the UK. For example, this measure was utilised by the US Agency of Healthcare and Research Quality as a part of the National Quality Measures Clearinghouse.

The CARE measure has also been used by researchers from South Africa for development of a short-term training programme for forensic interviewing, to assess oriental medicine practitioner empathy in the patient population in Hawaii, in teaching French GPs, and to identify the attitude of the Greek GPs towards frail patients in primary care (Mercer, 2009). Thus, as a measure of clinical consultation performance skills, the CARE instrument appears to be gaining momentum and demonstrating its value in a wide range of clinical domains as well as different health care settings.

4.3.5 The CARE questionnaire for medical students

The previous sections showed how the CARE measure was developed, validated, and is increasingly being used. This section discusses how the CARE measure may be of use with medical students.

The CARE questionnaire developed by Mercer and colleagues includes ten distinct statements and no free-text response options. As an overall global rating question has been shown to significantly correlate (Pearson's correlation = 0.922, $P > 0.01$) with answers to mean individual question scores (Murphy, 2009a), we decided to explore the utility of a simpler one question approach. Thus the questionnaire form in this study presented 11 items, including a global question to gather patients' views on the overall behaviour of student during the consultation.

A free-text option was included for patient's feedback from the second year of this study (Appendix 4) when we noticed that some patients included their comments on

their responses beside the scores. Such a free-text box was also added to the CARE questionnaire and used to gather patients' feedback by Mercer and colleagues in their study of capturing patients' views on communication with anaesthetists (Mercer et al., 2008b).

For the sake of practicality, a seven-digit code was created for purpose of coordination to identify the year, the student, and the sequence number of each questionnaire. For example, in the code number 09-028-15, 09 denotes that the questionnaire was distributed in 2009, 028 corresponds to student number 28, and 15 means that this was number 15 of the 35 questionnaires allocated to that student. In this way, the number of surveys for each student could be tracked, the confidentiality of student identity was preserved, and the anonymity of patient participants maintained. Patients were instructed not write their names on the questionnaires, which were anonymous. For the purpose of this study the CARE questionnaire form was printed on two-sided A4 paper. One side was dedicated to questions while the reverse side page contained the Patient Instructions (Appendix 5). This served six purposes:

- to describe the study and its purpose, in lay terms;
- to inform patients that the questionnaire was anonymous;
- to inform patients that completion of the questionnaire represented their consent for inclusion in the study;
- to guide patients to select 'Does not apply' where relevant rather than leave blank answers;
- to guide patients about completing the questionnaire and to return it to reception in sealed envelope; and
- to provide contact information in the event of questions or concerns.

4.4 Outline of the Dundee Medical School programme

Around 160 students are admitted each academic year and are involved in Dundee's five-year curriculum. Dundee Medical School utilises a spiral curriculum which is problem oriented and outcome based. Other features of the programme are early clinical contact across years 1–3, two years of clinical attachments (year 4 and 5), and a six-week elective period between years 4 and 5. The Dundee curriculum has been designed to meet the changes in healthcare services, public expectations of a doctor, and modern teaching methods and assessment approaches. Also, the curriculum attempts to match the students' needs by adopting a student-centred approach (University of Dundee, 2010) and meets the expectations set out in the *Tomorrow's Doctors* document.

Students in the first three years are introduced to both basic sciences and clinical practice within an integrated learning programme. The student progresses through a systems-based programme from middle of year 1 to the end of year 3 during which they learn basic clinical and consultation skills in both simulated and ward environments. At the end of year 3, there is a transition course to help prepare students for the clinical attachments in the next two years. This includes a short GP placement to consolidate students' understanding of general practice.

In the fourth and fifth years, the course provides the student with many clinical learning opportunities. Year 4 students experience clinical practice in hospitals and primary care context through ten 4-week rotations, with their learning anchored around over 100 'core clinical problems' to ensure breadth of study. One of these blocks the student spends in a general practice in urban areas such as Dundee, Perth, or other nearby towns. The students learn to consult patients under supervision initially but are expected to take a lead towards the end of their attachment. Tutors are requested to

ensure every student has opportunity to conduct at least 20 'solo' consultations, meaning they must consult independently up to the point they feel require assistance. The tutor will always review the patient and help conclude the interaction. Students are not allowed to complete consultations independently.

The final preparation for practice as a foundation doctor takes place during the fifth year of medical school. This is obtained by a combination of two 'Foundation Apprenticeships' and five additional one-month blocks (two theme-based, two hospital clinical and one in primary care). The GP block is generally in a rural area and the first two days involve core teaching sessions which cover immediate care, prescribing, and consultation. Some fifth-year students opt for an extended placement of eight or 12 weeks in general practice which runs as two groups throughout the academic year. These students also have a two-day debrief and an Objective Structured Clinical Examination at the end of their placement. The GP placement assessment process includes a mini-CEX, a case-based discussion, a professional attributes form and the tutor's final assessment and feedback. Tutors are again expected to ensure a minimum of 20 'solo' consultations are performed by each student within a standard one-month block, with those extending expected to do so during the first month and then continue doing so regularly thereafter. Therefore Dundee Medical School students are a promising group for study because they will conduct ample consultations (up to 40), can be afforded a settling-in period, and are more likely to be interested in general practice and therefore, the consultation process.

4.5 Participants characteristics and instruction forms

4.5.1 *Students*

Fifth-year medical students at the University of Dundee were asked to participate in this study during the academic years 2009–2010, 2010–2011, and 2011–2012. Students conducting independent consultations with patients during two- and three-month GP attachments were asked to participate. In the first year (2009–2010), this was entirely voluntary, while in the next two academic years, participation was a course requirement. One-month GP attachments were considered too brief to complete the required number of forms. Students from hospital rotation were not included in the study, as the method of distributing and collecting the questionnaires would be too complex and unreliable.

4.5.2 *Patients*

Consecutive patients (consulting primarily with the student) were sought to participate. We were aiming for 25–30 returned forms per student. In agreement with the University Research Ethics Committee, exclusion criteria included an inability to read and understand the questionnaire, and when the student or supervising GP felt it would be inappropriate (such as with critical illness or a distressing diagnosis). Patients below the age of 16 were excluded or the questionnaire was completed by accompanying adults. Patients with visual impairments who were otherwise eligible could have a companion assist. Forms could not be completed by the student or any staff member in the practice, in case this led to more favourable responses.

4.5.3 Teaching staff, tutors, and practice staff

The GP block director was informed about the study method during the preparatory phase. Issues concerning the process were discussed in detail to ensure a suitable process was devised. The block director informed the tutors about the study, in year 1 via email, and thereafter, via staff development sessions. This allowed GP tutors to integrate the questionnaire with other teaching or support they provided for the student. Practice managers were also alerted about the study and asked to assist students gathering forms.

Instruction forms were prepared and distributed to GPs and practices via email or posted to others along with any other material from the course director. This was done within the first two days of the GP block during the introductory session. The instruction form gave a full explanation of the study and asked that practice staff should ideally hand the questionnaire forms to the patients and then gather them to minimise bias. However, when this was not possible, either the student or the tutor distributed the forms. Ideally, forms were gathered in reception and returned to the researcher once at least 25 had been returned. The last two paragraphs of the instructions contained two important points:

- everyone was advised not to assist the patients in completing the questionnaire (i.e. responses were to be collected from patients in a truly anonymous fashion; and
- the students and tutors were reminded that this study was for assessing the appropriateness of the CARE questionnaire and was not to be used to assess the medical students at this stage.

Tutors were alerted in advance by the medical school office about the study and also received a copy of the CARE questionnaire to allow them to integrate its content with any other teaching or support they may provide for the student.

4.6 Data collection

4.6.1 *Student recruitment (introductory briefing)*

In the Dundee Medical School, final-year, medical students have options on both the length of GP attachment (one-, two-, or three-month attachments) and the location. Only those who completed the extended blocks (two- and three-month GP attachments) were invited/expected to participate in this study; they formed nearly 20% of all fifth-year students, representing approximately 25–45 students per year.

Students undertaking either two- or three-month attachments attended an introductory course over two days. In the first academic year of the study (2009–2010), the principle investigator was allocated 30 minutes of the introductory session to explain the CARE study and questionnaire distribution method to students. The students of each group were asked to participate and informed about the purpose of the study. They were reassured that this assessment tool would not be used to evaluate their performance at this stage of the study and the measure validity was not yet proved with students. The GP block director took this task during the second and third academic years of the study period as the CARE measure became a course requirement.

The students were provided with a large envelope (freepost, directed to the principle investigator) with 35 questionnaire forms and a same number of small blank envelopes. In addition, copies of instructions to students, tutors, and administration staff were included in the packet. Students were able to review each item in the packet before leaving, so at the end of the presentation, students had the materials required, had reviewed the instructions, and had been provided with the opportunity to ask questions.

4.6.2 Data management

At the practices, students handed practice managers the packet. The important thing sought was returning at least 25 CARE questionnaires per student commencing from the third week of the GP block to allow the student to settle-in and familiarise themselves with the setting. When it was impossible to distribute forms via practice staff, such as in practices which depend on touch-screen booking, then either the tutor or the student took responsibility for distributing the questionnaires at the end of the consultation. The students were emailed by the principle investigator just before the third week of the block and reminded to ask practice staff to start distributing questionnaires. In addition, two weeks before the end of the block, the researcher contacted practice managers and reminded them about forwarding the completed questionnaires. The completed questionnaire forms were gathered, and data was double entered manually onto a spread sheet, and checked for 'allowed characters'. Data was stored in a secure database and could only be accessed by the principle investigator.

4.6.3 Gathering students' views

The process of gathering students' views on usefulness and acceptability of the CARE measure during their clinical attachment period was conducted in the debriefing session at the end of each GP block. These sessions were led by the block director with the principle investigator in attendance. Each student was invited to give his/her views verbally which were summarised by the principle investigator and discussed with the supervisors. Suggestions and decisions were important to the next stage so that possible changes in the study plan or the method of data collection might be considered in the following groups.

Students' views and ideas were gathered to highlight the practical obstacles for medical students, especially for those who participated during the academic year 2009–2010. At the first year of the present study, most of the students were concerned that patients' ratings might affect their final scores, as they had no idea how much patients will be satisfied with their performance. This trend led students to be apprehensive about CARE, although these views were changed in the subsequent years, when students became more confident with patients' ratings after early figures revealed an encouraging high average. A formal assessment of the face validity of CARE from students' perspective was not conducted due to limited time and a sense that students are overloaded with feedback requests, including those required for their GP placement.

4.6.4 Gathering and utilising patients' comments

The CARE measure used during the first year of the study had no patient comment box. However, a considerable number of patients did comment on their consultation using the narrow space of the questionnaire paper. This led to a change in design, with addition of a box for free text added at the bottom of the form (Appendix 4).

The patients' comments for all students were collected and classified into two main categories of comments: 'positive' feedback and 'negative' feedback. Then the feedback was sub-categorised to specific or non-specific comments. The positive category of patients' comments reflected their satisfaction in general non-specific words and contained encouraging specific remarks for students. In contrast, the negative category of patients' comments reflected what patients pointed out concerning students' weaknesses during consultation. This was considered sufficient to

assess the added value of including the free-text box and allowed collation of these comments in more specific subcategories presented in detail within the Results Chapter (Chapter 5).

4.6.5 Provision of feedback to students

All students for whom at least 25 forms received were given feedback on their results by the principle investigator via an email during week seven of the block, prior to student appraisal. The CARE results' email sent to students (Appendix 6) thanked them for participation in the study and provided students with their mean scores for each question in form of small table. In the first year of the study, students who participated were sent feedback in the form of simple scores, as they could use these with the descriptors. From the second year of the study, the students had the option to compare their results with the norms obtained from former students; the norms were provided in a table attached to the email (Appendix 7). The mean and confidence intervals in that table were updated for each new group as more data was received. The students who did not complete the minimum needed number of CARE questionnaire forms (25 forms) were notified that their means might not be reliable enough to compare. When relevant, students were also sent patients' comments related solely to them. The students were instructed that they could decide to share the feedback with their tutors or not, as they wished.

4.7 Statistical analysis

Descriptive statistics was performed initially. The students' sample and their response rate were important to show the feasibility of this measure within this setting. The reasons for non-participation have value and were analysed in order to improve the methodology and address practical problems during the data collection. To avoid introducing bias, the number of missing and 'Does not apply' replies was limited to a maximum of two per questionnaire. CARE forms with more than two inappropriate values were excluded. This strategy was suggested by the RCGP Scotland (2003) and confirmed by Mercer et.al (2005). Although there are many ways to deal with missing data (Schafer and Graham, 2002), it has previously been shown that this method gives similar total scores compared with other approaches such as excluding questionnaires with any non-valued data, and has the advantage of optimising sample size. This method was applied in different settings of Mercer's and his colleagues' research (Mercer and Murphy, 2008, Mercer et al., 2008b, Mercer et al., 2005b, Mercer et al., 2008a) and was used in the Chinese version of CARE in primary settings (Mercer et al., 2011a).

This was followed by the descriptive analysis stage in which data processing and analysis were performed using the Statistical Package for Social Science (SPSS) version 20 (IBM Corp, 2011). The mean average and standard deviation of CARE measure scores overall and at item level was calculated for each student. The descriptive statistics provide summary variables such as mean, median, standard deviation, and confidence intervals (95%). The Student means from this study and the GP means from Mercer's study were compared using an unpaired *t*-test. The CARE measure evaluates patient empathy as an interval dependent variable (i.e. measured on a continuous rating scale of empathy behaviour that can give any mean value within a

given score range by a cohort of patients). Thus, as empathy was not treated as an ordinal variable, it required parametric methods for analysis.

G-theory was used for testing the reliability of the CARE measure as a performance assessment tool because this is an optimal means of addressing multiple sources of variance, as is typically the case with tools such as the CARE measure (Brennan, 2001) (Streiner and Norman, 2008). G-theory considers error in measurement related to each factor (facet) or variance and estimates the how much error resulted from that factor and/or from the interaction between factors. An additional advantage of using G-theory in this study is that one can mathematically determine how many patients are needed to provide a reliable score. G-study was also required for this study to calculate the extent of agreement within patients (raters) for each student (Streiner and Norman, 2008).

According to G-theory (as explained in Section 2.7.7 and Table 2.2), the forms of reliability in this study were internal consistency (question), inter-rater (patient) reliability, and overall reliability (ability of the CARE measure to differentiate between students). This approach considered students as the subjects of measurement and the other factors (raters and items) as sources of measurement error (Table 4.1).

Table 4.1 Types of reliability and facets of the CARE measure for medical students.

Type of Reliability	Facet		
	Differentiation	Generalisation	Fixed
Internal consistency	Student	Item (question)	Rater (patient)
Inter-rater	Student	Rater (patient)	Item (question)
Overall	Student	Item (question) Rater (patient)	N/A

The reliability (inter-rater or patient, inter-item, and overall reliability of test) was assessed using G-theory by using the GENOVA Suite of computer programmes (urGENOVA software; (Brennan, 2011); (Streiner and Norman, 2008). In addition, a D-study was conducted to determine the effect of changing the number of observations to assess the minimum number of raters (patients) per student (15, 20, 25, 30, 35, 40, ...).

The importance of assessing validity was explained in more detail in the chapter containing the literature review (Section 2.7.3). In fact, any assessment should have validity beside the reliability to back up the claims that the tool is a sound measure. The validity of the CARE measure, like any other measure, depends on acceptable reliability (Streiner and Norman, 2008). However, this is not sufficient; it must also show evidence of multiple sources of validity.

The CARE measure's concurrent validity was tested by correlation with other well-established measures of the related constructs that have been previously validated. The closest assessments to the CARE measure for fifth-year medical students in Dundee Medical School are the GP Objective Structured Clinical Examination (year 5), the

year 5 GP block tutor mean, and the mini-CEX (year 5). Validity was assessed by estimating association between the CARE measure scores and the scores of these tools for the participating students.

The OSCE has been developed by Harden and colleagues for the assessment of the clinical competence of medical students (Harden et al., 1975). It has been proved to be a valid and reliable way of assessing clinical skills in many different health professionals for surgical residents (Cohen et al., 1990), internal medicine residents (Petrusa et al., 1990), and dental students (Brown et al., 1999a) and has a Cronbach's $\alpha = 0.68$ with high inter and intra-domain correlations. However the consultation component of clinical practice carries a variable weight within different Objective Structured Clinical Examination formats and an assessment heavily oriented towards this element would be required for comparison with CARE.

The year 5 GP block tutor mean is an important part of phase 3 GP Attachment assessment process. This process consisted of a mini-clinical evaluation exercise, case-based discussion, and an evaluation of professional attributes, in addition to the tutor's final assessment and feedback. The students are expected to complete a minimum of two mini-clinical evaluation exercises and two case-based discussions, in an extended attachment. At the end of each attachment, the tutor completes one professional attributes form and gives a grade for each of the 12 Dundee outcomes on the final assessment and feedback form based on their own observations, the observations of others in the primary care team, and an evaluation of the student's performance using the assessment tools.

The mini-CEX was first developed in the US. It is used by an observer to evaluate the learners' clinical skills in addition to physical examination and clinical judgment in

various clinical settings with diverse patients' problems in 15–20 minute encounters, with immediate feedback. This clinical evaluation exercise was showed to be valid and reliable (Cronbach's $\alpha = 0.90$; (Durning et al., 2002); (Norcini et al., 2003). The University of Dundee Medical School has been piloting use of the mini-CEX (Appendix 8) as part of final-year teaching and assessment (Dundee Report GMC, 2009). Each fifth-year student is required to undertake at least four mini-CEX encounters in the GP block.

4.8 Summary

This chapter has outlined the methodological approach adopted for this study. The studies' strengths and limitations are explored in more depth in Chapter 6. The aim of the study was to assess if the CARE measure can offer a reliable, valid, and practical means of providing patient feedback on the consultation and humanistic skills of medical students in the clinical phase of their training. The research design, data collection, analytic approach, and ethical considerations were outlined and demonstrated carefully. This chapter also described the changes made to the original form of the CARE measure and the reasons behind that. The student and patient recruitment and selection methods were presented as well as the process of data collection. Finally, the statistical analysis plan was set out thoroughly to analyse the findings, which are explored in the next chapter.

5 Results of the CARE Measure Utility

5.1 Introduction

This thesis chapter covers the research into aspects of the utility of the CARE questionnaire with particular attention given to its reliability, acceptability, and feasibility when applied in the medical student context.

The first section will give an overview of the descriptive statistics of the relevant quantitative variables. The study quantitative results into the reliability of the CARE measure will then be discussed as evaluated using G-theory. In addition, the correlations between CARE measure and other assessment instruments will be explored to evaluate concurrent validity. The latter part of this chapter deals with the qualitative aspects of the study and focuses on the acceptability and educational impact of the CARE measure by reflecting on the themes of patients' comments on their experiences of their consultations with the medical students. The results of a mapping exercise to explore the testing of students' curriculum coverage of needed GMC professional criteria and attributes will be presented in a separate chapter.

5.2 Descriptive statistics

5.2.1 Student sample and response rate

A total of 109 year-five Dundee Medical School students were engaged in two- or three-month GP attachments during the study period, and all were invited to participate in the study. The student population was distributed over the three academic years (2009–2010, 2010–2011, and 2011–2012) of the study, with 27, 37, and 45 students, respectively, from each year (Table 5.1). In the first year of the study, during the

academic year 2009–2010, a disappointing response rate of 55.6% was obtained. However, the response rate improved in subsequent years: the response rate of students during the academic year 2010–2011 was 86.5%, and the response rate reached its highest level (95.6%) during the academic year 2011–2012 (Table 5.1).

Table 5.1: The distribution of the study sample according to the students' response rate.

Academic year	No. of students who participated	No. of students who did not participate	Total no. of students	Students' response rate (%)
2009–2010	15	12	27	55.6
2010–2011	32	5	37	86.5
2011–2012	43	2	45	95.6
Total	90	19	109	82.6

As can be seen from Table 5.2 below, only 66 of the participating students returned the requested number of questionnaire (25 or more) forms. The remaining participants (40%) were distributed between those who did not complete the required number of forms (22%) and those who did not collect any forms (17.4%).

The total number of CARE questionnaires collected throughout the period of study was 2,145. As Table 5.2 shows, 1,808 forms were collected by 66 students who succeeded in returning the specified minimum of 25, with an average of 27 forms per student. The remaining 337 forms were collected by 24 students who failed to collect the minimum adequate number of questionnaire forms. Eight of the students who

collected ≤ 24 CARE questionnaire form collected 20–24 forms, 11 students collected between 10–19 forms, and five students collected less than nine forms.

Table 5.2: The distribution of participants according to completion of minimum required number of questionnaire forms.

Academic year	Students who completed ≥ 25 forms		Students who completed ≤ 24 forms		Total no. of participants in each academic year	
	No.	%	No.	%	No.	%
2009–2010	11	73.3	4	26.7	15	16.7
2010–2011	27	84.4	5	15.6	32	35.6
2011–2012	28	65.1	15	34.9	43	47.8
Total	66	73.3	24	26.7	90	100

5.2.2 Reasons for non-participating

These were related either to the GP or the students themselves and can be summarised as follows:

- GP Tutor
 - 1- GPs did not allow the student to undertake the task.
 - 2- The GP declined to participate in the study.
- Student
 - 1- The student was worried about the scores.
 - 2- The student did not wish to take part in the CARE study.
 - 3- The student started the attachment too late.

5.2.3 Reasons for not completing the minimum number of questionnaires

At the end of each GP training block, students were interviewed and asked about their experience of collecting the CARE questionnaire forms. Students who failed to complete the needed number of CARE questionnaires cited a number of reasons for difficulties at a practice level. These included reasons related to the practice they in which they were being trained and processes around handling of forms by reception staff. Difficulties were also experienced at a patient level. These barriers prevented students from collecting the specified 25 CARE forms. Specific difficulties included:

- Practice issues
 - 1- Small practices (difficulties in managing the number of patients in time frame).
 - 2- The number of surgeries number was not enough (e.g., some had been cancelled due to snow).
- Reception issues
 - 1- The reception staff were too busy and forgot to hand out forms.
 - 2- The reception uses a touch-screen check-in process.
 - 3- Lack of staff motivation in distributing the questionnaire.
- Patients' issues
 - 1- Patient left without returning the questionnaire.
 - 2- The patient brought back the forms at the following appointment.
 - 3- The patient didn't have his/her reading glasses.
 - 4- There were language limitations.
 - 5- The patient did not have enough time to fill out the form.

In addition, in spite of explicit instructions to patients and practice staff that envelopes should be sealed by patients and handed to reception staff, a small number of the forms received were in opened envelopes and others not in envelopes at all, which increases the chances of the submission of fabricated forms by students. Therefore, these forms were also excluded from the study.

5.2.4 *Missing data*

In order to test the applicability of individual CARE measure items in the medical student training context, both missing and ‘not applicable’ responses were counted. It was found that from the total number of CARE forms collected (2,145), there were only 799 (3.4%) non-valued items (‘Missed’ and ‘Does Not Apply’) among the total number of items (23,595). Most of the non-valued items were ‘Does not apply’ 749 (93.7%), while there were only 50 (6.3%) missing values.

As can be seen from Table 5.3 below, CARE questions 1 to 7 elicited very few missed values or ‘Does Not Apply’ responses. Likewise, similar high rates of completion were found for the questionnaire’s overall global score (Q11). The percentage of missed values increased slightly for question 8, which evaluated students’ explanations to patients. The highest percentages of ‘Does not apply’ and missing values were 40.6% and 36.8% for questions 9 and 10, respectively.

Table 5.3: Numbers and percentages of non-valued data per each question of the CARE measure.

CARE questions	'Does not apply'		Missing		Total of non-valued data	
	No.	%	No.	%	No.	%
1. Making you feel ease?	0	0	0	0.0	0	0.0
2. Letting you tell 'Your' story?	4	0.5	1	2.0	5	0.6
3. Really listening?	1	0.1	1	2.0	2	0.3
4. Being interested in you as a whole person?	25	3.3	3	6.0	28	3.5
5. Fully understanding your concerns?	20	2.7	0	0.0	20	2.5
6. Showing care and compassion?	11	1.5	2	4.0	13	1.6
7. Being positive?	26	3.5	8	16.0	34	4.3
8. Explaining things clearly?	63	8.4	7	14.0	70	8.8
9. Helping you to take control?	317	42.3	7	14.0	324	40.6
10. Making a plan of action with you?	280	37.4	14	28.0	294	36.8
11. Overall behaviour during consultation?	2	0.3	7	14.0	9	1.1
Total	749	93.7	50	6.3	799	100

5.3 Descriptive analysis

5.3.1 Mean and standard deviation of CARE measure scores

The mean scores from the CARE measure were calculated by using forms which had no missing values and/or 'Does not apply' responses (1,713 included responses out of 2,145 patients) and so excluded 20% of patient respondents. This calculation gave an overall CARE question mean score of 4.57 (standard deviation = 0.19, median = 4.58). The mean of each question was then calculated and the non-valued data ('Missed' and 'Does Not Apply') were replaced by the average column for each question for all patients rated all students to allow statistical software analysis.

Figure 5.1 below shows the average scores for each question. Most students put their patients at ease and let the patient tell his/her story, and most students listened to their patients effectively. Clinical management questions (9 and 10) exhibited both a lower response rate and a drop in the mean for all questions. The highest average score was recorded for the overall behaviour of the student during consultation.

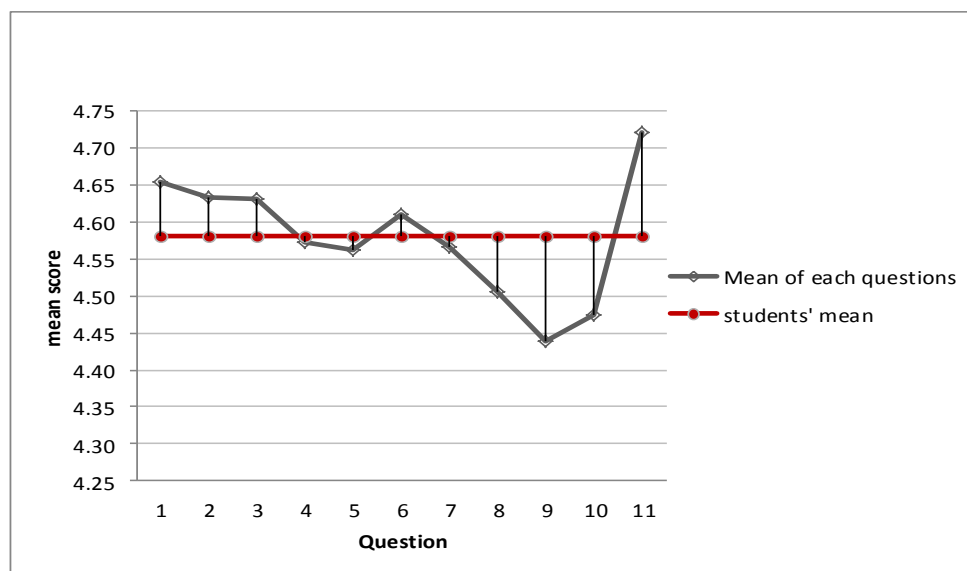


Figure 5.1: The average scores of each question of CARE questionnaire versus the overall mean.

5.3.2 Distribution of students' means and confidence intervals

The majority (61%) of the patients' scores were excellent, and less than 5% of the whole grades were considered by provided descriptors of scores as fair or poor. The mean scores and their 95% confidence intervals are shown in Table 5.4 below. Most students were judged as 'excellent'. Students with a mean score ≤ 4.14 can be considered to be significantly below average, and any student with a mean score ≥ 4.93 can be considered as significantly above average. The finding of the study revealed that of the 66 students, two students were below average and another two above average.

Table 5.4: The CARE measure students' mean scores, standard deviation (SD), and 95% confidence intervals.*

CARE Questions	Means	SD	95% Confidence Intervals	
			Lower	Upper
1. Making you feel ease?	4.61	0.17	4.29	4.97
2. Letting you tell 'Your' story?	4.62	0.18	4.24	4.97
3. Really listening?	4.62	0.18	4.22	4.95
4. Being interested in you as a whole person?	4.56	0.23	4.65	4.99
5. Fully understanding your concerns?	4.55	0.22	4.10	4.94
6. Showing care and compassion?	4.61	0.19	4.19	4.96
7. Being positive?	4.56	0.22	4.08	4.98
8. Explaining things clearly?	4.51	0.22	4.01	4.93
9. Helping you to take control?	4.43	0.26	3.88	4.90
10. Making a plan of action with you?	4.47	0.28	3.86	4.98
11. Overall behaviour during consultation?	4.71	0.18	4.31	5.00
Total Averages	4.57	0.19	4.14	4.93

* n = 66 students

5.3.3 *Students' mean scores distribution*

The overall mean of students' CARE scores with ≥ 25 forms was 4.57 (standard deviation = 0.19, $n = 66$) with a median of 4.58. The histogram of the students' mean CARE measure scores was not distributed normally, as can be seen in Figure 5.2. In this study, the mean and median fell at different points, with the median value higher than that of the mean. Therefore, students' means were highly skewed negatively (-1.354) as most of the mean values were concentrated on the right of the overall mean with extreme mean values to the left. The peak of the distribution indicated that students' means were leptokurtic (4.691) (i.e. sharper than normal distribution) and most of the values of students' means were concentrated around the overall mean.

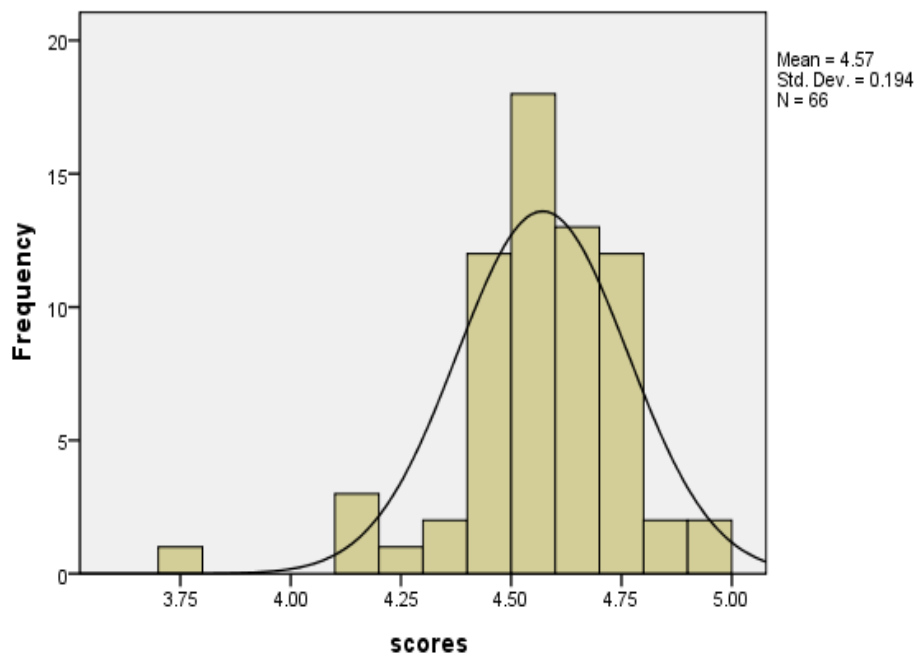
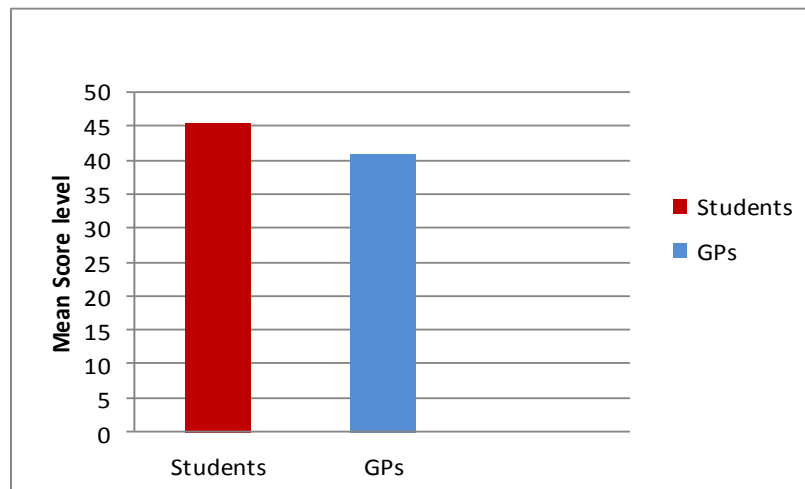


Figure 5.2: Students' mean CARE measure scores distribution (histogram of means)

5.3.4 *A comparison of student and physician CARE score outcomes*

Another method to quantify CARE measure scores used by Mercer et.al was to calculate the sum of question item scores. Here, the scores of question 11 (overall

behaviour) were excluded to match the number of questions for original CARE questionnaire; therefore, the total number of CARE forms with less than two missing values and/or 'Does not apply' was 2,061. In this case, the overall mean score was 45.6 (standard deviation = 5.32, $n = 2061$) with a median of 48.0 (range 20–50). By comparison, the total number of CARE forms for the GPs study was 2,734, with mean 40.8, standard deviation 8.8, and median 41 (Mercer et al., 2005b). The difference between the student mean and the GP mean is illustrated in Figure 5.3. The students' mean exceeded that for GPs and the difference is considered to be highly statistically significant ($t = 21.94$, $P < 0.001$). Also, the students' mean was higher than the means of doctors in secondary care (43.5) and those of anaesthetists (43.8).



$p \leq 0.001$

Figure 5.3: Difference between the student and GP mean CARE measure scores.

5.4 Qualitative results

This section of the thesis gives an overview of the feedback from patients to students. It shows the overall views of patients and the types of positive or negative comments made available to students for their reflection on their empathy skills. This capture of free-text comments was facilitated by the inclusion into the original CARE measure of a free-text box allowing patients to express their experiences by written comment. In addition, the section will explore the views of students and staff on the use of the CARE measure with medical students.

5.4.1 *Patients' feedback (Comments)*

A 264 (17.6%) patients out of a total number of 1,495 patients commented on their independent consultations with students (from the second year of the study 2010–2011 when this free-text option was added). Free-text comments were recorded for 55 students, with an average of four comments per student. Most of the patients' comments expressed their satisfaction with the consultation and could be divided into two main categories as shown in Table 5.5 below. The majority of patients' comments (83.7%) were positive or affirming, said that students were generally good or excellent, and provided suggestions to encourage their perceived high level of empathy skills during clinical consultations. The remaining comments (16.3%) were negative or corrective in nature and pointed to perceptions of students' inappropriate behaviours during the consultation.

The results of the present study revealed that most of the positive feedback interventions offered by patients were non-specific, whereas all the negative comments were specific. Overall, we found that approximately 48% of the patients' comments consisted of specific feedback.

Table 5.5: The main categories of patients' comments.

Main categories of patient comments	Number	Percentage
Positive Comments	221	83.7
Negative Comments	43	16.3
Total	264	100

5.4.1.1 Positive patient comments

Positive patient comments are distributed into two subcategories as illustrated in Table 5.6. Over a third of the positive category of patients' comments were specific comments. The non-specific comments comprised 62% of the positive type of patient comments. In general, the patient comments described students as 'helpful' or 'lovely', and other patients commented on the good listening skills of the students and praised students' empathic behaviours. Some of the patients wrote about their feelings towards the caring level of the students and satisfaction with consultations. Patients also commented on their expectations about students as doctors in the future.

Table 5.6: Specific and non-specific positive patients' comments.

Category of positive patients' comments	No.	%
Specific comments	84	38
Non-specific comments	137	62
Total	221	100

Some examples of these positive specific and non-specific comments from patients on consultations are as follows (the full list of positive comments attached as Appendix 9).

Specific patient comments:

- ‘He listened to my concerns and gave me appropriate advice very kindly.’
- ‘The student was very easy to talk to and reached to things in a positive manner making me feel comfortable to talk about even embarrassing things. I would suggest that she get as much experience as possible and sure she will make an excellent doctor.’
- ‘This student was excellent, and builds up a good rapport, taking time to listen.’
- ‘Spoke well and actually took the time to listen to everything. Had to say 10.10 for communication skills and making me feel at ease (thank you).’

Non-specific patient comments:

- ‘She will make an excellent doctor. She has it all now.’
- ‘Very relax consultation, excellent.’
- ‘I was quite happy with the consultation.’
- ‘Consultation was super.’

5.4.1.2 Negative patient comments

The percentage of negative comments in this main category was lower than that of positive comments. A perceived lack of confidence was the main specific sub-category, and this formed nearly the third of all critical patient comments. Seven comments criticised the student's approach to history taking or clinical examination. Ten comments considered aspects of professional behaviour of students during the consultation (Table 5.7).

Table 5.7: The specific categories of the negative patients' comments.

Specific Category	No.	%
Confidence	14	32.60
Communication	6	13.95
History and examination	7	16.30
Professionalism	10	23.20
Nervousness	6	13.95
Total	43	100.00

The full list of negative comments can be found in Appendix 10. Below are examples of negative patient comments:

- *'Possibly try to sound more confident in her diagnosis.'*
- *'I did feel she said sorry too frequently, while examining me even i thought i reassured her she wasn't hurting me. Otherwise she was excellent.'*
- *'Don't be scared, most patients don't bite.'*

- *‘Just one. Check that the patient is not slightly deaf, if so, speak more slowly.’*
- *‘She seemed slightly nervous, but I’m sure she’ll get over that.’*

Some patients found the students nervous, unclear in speaking with patients, or even that the students did not listen well to them. A few comments pointed some of the undesirable personal issues in students listed below:

- *‘Dress code could be smarter and more appropriate for the surgery, otherwise very satisfied.’*
- *‘Stop hitching up skirt during consultation, distracting.’*
- *‘Not to yawn three times during a consultation.’*

5.5 Quantitative results

5.5.1 *Reliability of the CARE measure with Students*

The reliability of CARE, that is, its ability to differentiate between students' levels in their patients assessed levels of empathy for a given number of patient questionnaires, is of key importance to understand. This will be dependent on the number of patient opinions collected. Increasing the number of opinions helps form a stable view for an individual student and minimises the potential for error. In addition, knowing the reliability offered by a given number of questionnaires helps decide the feasibility of an applied system as to whether the reliability from a 'possible' number of questionnaires is practical to collect and of a level fit enough for the purpose for which it is applied.

Table 5.8 illustrates the internal consistency, inter-rater (patient) reliability and overall reliability of CARE measure. Each is a number between zero and one, indicating the extent to which one can take account of error in scores due to the question (item) (internal consistency), the patient (inter-rater), or both of these sources combined (overall reliability). The first column in Table 5.8 indicates the reliability that could be achieved with different numbers of patients (raters) to help assess the feasibility of CARE as a tool to rate students. The highlighted cells mark the data of this study, and G-theory calculations showed that the internal consistency of the CARE measure was very high (0.97). With 25 patients per student, we can achieve an overall reliability coefficient of 0.74. The inter-rater reliability of the CARE measure was thus found to be high in terms of its ability to discriminate consistently between students.

The feasibility of the CARE instrument was investigated by performing D-studies on the data collected. D-studies allow the mathematical manipulation of the different

potential sources of variance from a study. This allows the exploration of the tested measure's different forms of reliability for any given set of CARE questions or number of completed ratings (patient questionnaires). Table 5.8 illustrates the reliabilities of CARE for between 10 and 40 patient questionnaires per student. These D-studies revealed that an acceptable reliability of 0.7 can be achieved with 20 questionnaires per student, and even with 15 forms, we still reach an overall reliability of (0.64) overall reliability (Table 5.8).

Table 5.8: The CARE measure reliabilities according to number of patients.

Tool	No. of raters (Patients)	Internal consistency	Inter-rater reliability	Overall reliability
CARE	10	0.96	0.56	0.55
	15	0.96	0.65	0.64
	20	0.97	0.72	0.70
	25	0.97	0.76	0.74
	30	0.97	0.79	0.77
	35	0.97	0.82	0.80
	40	0.97	0.83	0.81

5.5.2 Validity of the CARE measure with students

Table 5.9 presents the correlations between students' mean scores of the CARE measure and scores of other assessment tools that measure the same elements of consultation skills of student sample of the present study. The CARE measure was moderately, but significantly, correlated with other current measures of clinical

consultation aspects for year-five medical students, such as the GP OSCE, tutor assessment, and mini-CEX ($r = 0.325$, $r = 0.329$, $r = 0.371$, respectively; $P < 0.05$).

Correlations between the GP OSCE and both the mini-CEX and tutor assessments were a little stronger. These results will be discussed in the following chapter.

Table 5.9: Correlation of CARE students' mean with other assessments of same construct.

Assessment Tools		CARE students' means	Y5 Mini-CEX	Y5 GP OSCE end of block	Y5 GP block tutor mean
CARE students' means	Pearson correlation sig. (2-tailed)	1.000			
Y5 Mini-CEX	Pearson correlation sig. (2-tailed)	.371*	1.000		
		.016			
Y5 GP OSCE end of block	Pearson correlation sig. (2-tailed)	.325*	.414**	1.000	
		.035	.006		
Y5 GP block tutor mean	Pearson correlation sig. (2-tailed)	.329*	.840**	.540**	1.000
		.033	.000	.000	

*Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed)

5.6 Summary

The results of this study showed an increasing response rate by students, especially during the second and third years of the study period. The collected raw data contained a low percentage of non-valued patients' scores, and most of these were related to the questions 9 and 10 of the CARE questionnaire. The majority of students obtained excellent CARE measure mean scores, higher than postgraduates in GP settings. A reasonable number of patients commented on their consultation with student and were generally very positive. However, some comments were negative and reflected the weak areas in relational empathy of the students. The CARE questionnaire showed high internal reliability and effectively differentiates between students, provided at least 20–35 ratings per student can be gathered (inter-rater reliability = 0.7, 0.8). In addition, the CARE measure was significantly correlated with other current measures of clinical consultation aspects for year-five medical students such as GP OSCE, tutor assessment, and mini-CEX.

6 Discussion of the CARE Measure Utility

This chapter discusses the findings of the study, which set out to investigate whether the CARE measure offers a reliable, valid and practical means of providing patient assessments of consultation skills of medical students in the clinical phase of their training.

The CARE measure has been already been subjected to sufficient psychometric evaluation to be considered a useful measure for practicing clinicians. This measure has the advantage of being brief and understandable by patients (Mercer et al., 2004a, Mercer et al., 2005b) in different settings (Mercer et al., 2008b, Mercer and Murphy, 2008). This study extends previous work into the undergraduate arena and further supports the reliability and validity of the CARE measure with senior students.

This chapter first discusses how the CARE measure may be employed in the general practice context with medical students. Then the issues surrounding the CARE measure utilisation are discussed in the light of the study findings. Finally, the possible methodological weaknesses of this study will be highlighted and discussed.

6.1.1 The method of CARE measure utilisation for medical students

As noted in the methods chapter (Chapter 4), the distribution and collection of the forms was conducted in the third week of the attachment periods. This period was considered important so students could settle into their practice and was chosen to give them time to conduct some independent consultations before distributing questionnaires. It was thought that students would feel more comfortable and less stressed consulting with patients after a settlement period; also it was hoped that this

approach would reduce the effects of inexperience and therefore provide a more accurate picture of students' performance.

CARE form collection was concluded at the end of sixth week of the block, two weeks before the end at which time the GP tutor rated and the GP OSCE exam assessed their performance; this allowed the investigator to enter the data, analyse it, and give feedback to each student with patients' mean scores and comments. Therefore, all students were encouraged to send back the completed forms as soon as they reached the minimum number needed for this study.

This approach appeared to work well but is an approach that could only work in attachments of at least 6 weeks or some form of longitudinal arrangement. We did not assess staff or student views on the feasibility of gathering forms in shorter or ward attachments.

6.1.2 Effect of Student response rate on the CARE measure utilisation

To be a volunteer for new assessment measure is not pleasant thing, perhaps especially if you are a final-year medical student. Despite reassurance and clarifying the aim and non-summative approach of the study, the response rate of the students was frustrating initially before participation became compulsory in the following two years. Students were initially nervous about using the CARE questionnaire; but this did not mean it was an unfeasible measure, as it became accepted once familiarity grew in the second year when the students saw that their past year colleagues' CARE scores were generally excellent with encouraging patients' comments and provided something positive to include within their portfolios.

In addition, this mandatory approach adopted in this study was introduced because the faculty could see the educational benefits of and students' enthusiasm for feedback direct from real patients. However, it must be acknowledged that the initial reluctance encountered is something that others are also likely to meet when students are being asked to engage with a further and novel form of assessment. The educational benefits behind the above approach were under the aegis of 'assessment drives learning', although this relationship was called for more research and discussion and still under academic debate (Schuwirth and Van der Vleuten, 2004b, Van der Vleuten and Schuwirth, 2005). The relationship between assessment and learning remains a challenge for researchers, and this issue needs more studies on monitoring assessment and evaluating its effect on learners (Van der Vleuten et al., 2010).

6.1.3 Why use 25 CARE questionnaire forms?

It was planned that at least 25 or more CARE forms for each student would be collected; this was determined to be a feasible target based on discussions with the organisers of the GP block and also likely to be sufficient, given previous experience and data on the scales performance. It was also based on the likely number of independent consultations that the students would complete during the extended week block. According to the literature and based on the past few years' experience that GP block directors have, it was found that most of the practices where students might train could offer more than that number of teaching consultations during the specified period. In addition, the study findings supported this number, as the majority of the students who took part succeeded in collecting 25 forms, often before the six-week cut-off, and the average number of collected forms was actually higher. In addition, Mercer found that, using 25 patients per doctor, the CARE measure has the ability to

discriminate between anaesthetists with overall reliability of 0.74 (Mercer et al., 2008b). Accordingly, students who were undertaking one-month attachments were not invited to participate in the study, as the period was deemed too short to be useful, taking into consideration the need for a settling-in period and for data entry and analysis on the end of the block. However, depending upon the local arrangement, there may be alternative approaches that could make this possible (e.g. 10–15 forms could be required for purely formative purposes or additional responses could be gathered during other attachments).

The option of using fewer forms was supported by a D-study which indicated the number of forms per student required for an overall reliability level 0.7 would be 20, or 0.64 for 15 forms/student. These findings indicated that 25 forms (patients) per student may be needed and acceptably reliable for summative assessment but that 15–20 forms, potentially during one-month attachments, would also provide valuable information and in particular, be adequate to identify any particularly weak students for remedial intervention.

6.1.4 Reasons for not completing the minimum number of questionnaires

The source of information about reasons for not submitting the expected number of questionnaires was either from emails sent by students to the principle investigator or block organiser during the GP attachment or from the students themselves when we met them at the end of the block briefing session. Failure to collect 25 or more forms was attributed mainly to different issues related with the practice size, number of patients consulted (i.e. the student had less opportunity to consult independently), and new technology (touch-screen check-in) in reception which kept patients from direct

contact with reception staff. In addition, as mentioned in the methods chapter (Chapter 4), the GP attachment training of fifth-year medical school students in Dundee was conducted at practices located in the countryside, and some of these practices are small, sometimes with only one reception staff member. This was another reason that prevented students from completing the required number of CARE questionnaire for this study. Therefore, some adaptations were added to the methodology for distributing the CARE forms in specific situations where the student or tutor themselves (e.g. not the receptionist) handed questionnaires to patient. The practice issues reasons were considerably reduced as a result.

Changes made in methods to improve participation succeeded in lowering the percentage of non-participants and resulted in an elevated percentage of students who completed the minimum required number of forms. This is not considered to be a negative point at this stage, since we know that the majority of them completed at least 15–24 forms. Moreover, this might provide acceptable reliability in certain situations, as we will discuss in detail later on this chapter.

The finding of this study also revealed the insignificant differences according to completion of the required number of CARE forms per student (≥ 25 questionnaires). This does not support the hypothesis that weaker students collect fewer forms but indicates the cause as being related to the nature of reasons listed in the previous chapter, reasons which were out of the students' control. In addition, the majority of students collected and completed at least ten forms; this fact may support the efficiency of using that number of questionnaires per each student but only after taking into account the inter-rater reliability.

Form all the reasons mentioned above, it seems clear that in practice, students need to be pushed to seek responses. We were unable to explore the reasons why some patients elected not to complete the feedback or accurately measure the scale of this.

6.1.5 Reporting and handling missing data

The relevance of the individual CARE measure questions to the current consultation with patients is supported by the low percentage of missing values or ‘not applicable’ responses. Any of the CARE measure items marked as ‘not applicable’ or not marked indicated that those items were not of interest for the patients (Mercer et al., 2005b).

In this study, the overall low percentage (3.4%) of missing values and ‘Does not apply’ responses suggests the CARE questionnaire items were generally relevant to patients. This was expected for the present study also based on the previous studies of use of the CARE measure in primary care settings (Mercer et al., 2005b); (Mercer et al., 2011a). However, this was still of concern to us to avoid bias in findings and in the interpretation. A number of approaches exist for reporting and handling missing data (Schlomer et al., 2010). The finding of this study included different values of students’ mean score, taking into account the status of inclusion or exclusion of the non-valued data. In practice, as has been suggested previously by RCGP and accepted by Mercer in his study in general practice (Mercer et al., 2005b), scoring the CARE measure should include up to two invalid responses per measure.

The differences revealed the insignificant effect of missing values and ‘Does not apply’ responses of patients on the students’ means. This might be related to the ways of replacing the non-valued data with the average of the column (e.g. mean substitution); this supported the validity of that method of calculation and did not

affect the measure of central tendency. The percentage of missing data in this study did not exceed the limit that statistical analysis indicates is likely to be bias. However, experts have not reached a consensus regarding the percentage of missing data that becomes problematic (Schafer, 1999); (Bennett, 2001); (Peng et al., 2006).

Descriptive analysis revealed that the number of non-response values (missing values) were far fewer than 'Does not apply' responses, which suggests that there were some clinical cases that did not mesh well with the content of the students consultation (in particular, as is noted in Table 5.3, the percentage of non-valued data for both questions 9 and 10 which dealt with the management aspect of the interaction). As the consultant in this study was the student under training, it seems highly likely that some specific management plans were addressed by the GP tutor; according to the instructions, patients were to choose 'Does not apply' for questions they felt were not related to their consultation with the student. Even so, similar results were found by Mercer et.al. (2005), questions 9 and 10 of the CARE questionnaire had the highest percentages (10.8% and 14.9%, respectively) of 'not applicable' responses in their study (Mercer et al., 2005b). This is also similar to the finding of the cross-sectional study conducted in Hong Kong, which revealed that the Chinese version of the CARE measure showed a high proportion of patients with not applicable responses for question 10 (21.5%, with an average of 5.7% across all ten questions; (Mercer et al., 2011a). Our findings suggest both questions 9 and 10 may to be less relevant to students' consultations, which may raise option of deleting them and shortening the form. Alternatively, it could be seen as highlighting an area of concern, that undergraduates rarely get experience in participating in this final aspect of the consultation process, even in the highly supervised context of general practice. A

CARE questionnaire with only eight items could be more valid and reliable with medical students.

To reduce the problem of missing or ‘not applicable’ values for the CARE measure in the future, one option may be providing the CARE questionnaire online, as this method can encourage patients to respond to every item. This method of CARE questionnaire data managing could also facilitate faster analysis and feedback but could exclude less technically confident or able respondents and require a sophisticated IT infrastructure for it to operate smoothly. From all the above mentioned solutions together with the strategy of ignoring CARE forms with more than two missed or ‘not applicable’ values, which was applied to control the missed data of this study, in addition to replacing those values of both missing and/or ‘Does not apply’ with the average of that column for each question, we can say that this issue was managed effectively at the end.

6.1.6 Interpreting students’ scores

Generally, the students’ average scores were high, which suggests patients’ had positive feelings about their interactions. Patients do report altruistic reasons for allowing students them to participate in their caring, even if they are sometimes uncomfortable with their presence (Passaperuma et al., 2008). Although scores may be inflated, students can still gain useful information by comparing the categories in which their highest and lowest scores occurred.

The patients who rated the student were only those who wished to participate, perhaps less seriously ill and in less distressing situation. The healthier patient likes his/her doctor more and is more satisfied with care than the patient with worse health (Hall et

al., 2002). This phenomenon was expected to be similar to some extent in case of medical students, as most of their consultations were likely to have been with healthier and relatively more uncomplicated cases (as the practice staff and tutors instructed at the beginning of the attachment). The high scores achieved by students were confirmed by positive patients' comments.

As expected, students with low CARE measure scores also received lower scores on other communication skills' assessments such as GP OSCE and Mini-CEX. Interestingly, a study by Mercer and Howie has suggested that doctors who are rated lower by patients on the CARE measure have lower scores on aspects of self-rated morale and patient centredness (Mercer and Howie, 2006) This observation raises the concern that these low-scoring students may have additional broader issues to address.

6.1.7 Student mean scores

The mean levels for CARE in this study suggest that most students were judged as excellent. This result was expected from patients as they were rating junior doctors or medical students, and is supported by previous research conducted by Coleman and Murraya (Coleman and Murraya, 2002) as they found that patients liked to be involved in community-based teaching and they wished to make valuable contribution. The means and the 95% confidence interval levels can be used as cut-points but may be interpreted in two ways: they can be used either to define 'average' (4.54), 'above average' (4.83), or 'below average' (4.14) scores or to identify extremely good or extremely poor performers, according to norm-referenced results. Thus, the CARE measure can be used to reliably identify the weaker performers. However, when compared against the criteria patients were given; even the weakest students are

typically rated as ‘good’ (mean CARE score = 3.75/5). Considering how these students are rated within observational assessments such as OSCE and Mini-CEX, it appears that patients’ views in this context are overly favourable and that peer referencing is more appropriate.

Students assessed using CARE typically emerged with universally encouraging scores. Some had specific weak areas and a small proportion could be identified as universally weak. As an educational tool, CARE can therefore be used diagnostically to some extent, although care is needed if less than 15 responses in each area (including management and closure which may be less frequently scored) are obtained. The nature of remedial intervention was not the focus of this thesis.

6.1.8 *Student means versus GP means*

There are many factors associated with or that influence the CARE measure scores but it was clear that students’ scores are generally higher than those received for practicing doctors. Some of the factors that Mercer and his colleagues stated in their research (Mercer et al., 2005b, Mercer and Murphy, 2008) may also influence the students’ scores, such as ‘knowing the doctor well’. Higher scores might be expected by patients when they rate the student if he/she is under training of his/her preferred GP. Another factor was the consultation length, which was found to positively influence patients’ scores; as the time given for students is longer than that for everyday consultations with GPs, patients may have been more relaxed in telling their whole story. All of these factors may have provided the students with opportunity to establish a good relationship with the patient. In addition, as mentioned above the patients also appear to be viewing the students kindly (as compared with other exam data) so it is unlikely

that these higher scores are truly comparable. Though it may be interesting to consider how much consultation scores improve in practice when GPs are given additional time.

6.1.9 Reliability of the CARE measure for students

The aim of the present study was to determine the effectiveness of CARE measure in discriminating between students in primary care settings. We found that the CARE measure discriminates well between students (as well as doctors) if 25 patients complete the questionnaire per student, although 35 or more were required to give an ideal reliability of over 0.8. The need to provide only 25 CARE patient questionnaires with an overall reliability of 0.74 should not prove exhausting to meet and offers a feasible method for assessing students' consultation skills. The first trial of the CARE measure to discriminate performance with GP registrars in the context of general practice was conducted by Murphy (2009). He found that appropriate levels of inter-rater reliability (>0.8) could be achieved with 41 patients, with 24 patients required to achieve >0.7 ; this finding which was very close to the results with students found by Murphy (Murphy, 2009a). The present study supported the ability of the CARE measure to discriminate in performance between homogenous group (students) who were of the same age and experience. This was also demonstrated by Murphy in his research when the CARE measure was applied on GP registrars (Murphy, 2009a).

The inter-rater reliability of CARE measure was examined in this study to determine the number of questionnaires required per student to attain a reliable score on each student. The test-retest reliability of the CARE measure could not be calculated because patients only made one rating on one occasion. The number of patients required in the UK studies of the original CARE measure was between 40–50 patients

per doctor to attain a reliable score in both primary and secondary settings (Mercer et al., 2008b, Mercer et al., 2005b, Mercer and Murphy, 2008), and a fewer number (around 30 patients) in the case of the Chinese version of CARE (Mercer et al., 2011a).

Sources of variability were found with the CARE measure as it is a measure of professional performance, although the CARE questionnaire has a reasonable sample of items which produced stable and reliable scores (Van der Vleuten, 1996). In addition, this measure has other characteristics that support and increase its reliability, such as its use of different contexts and occasions, and the use of several judges (patients) is an important factor in minimising inter-rater bias and gaining a fair impression of the students' behaviour, leading to reliability enhancement of the measure (Crossley et al., 2002a). In this study, the potential sources of variability (raters or patients) were adequately sampled, and their effects on the precision of the measurement were diminished.

Improving the reliability of the CARE measure is more easily done than in case of other assessment tools. For example, the reliability of the OSCE can be improved by widening the sample of cases and standardisation of cases, but this increases the expense and labour involved. When the OSCE used for assessing communication skills, 37 different scenarios for a reliable empathy assessment were needed (Colliver et al., 1998). However, a patient perception questionnaire like the CARE measure seemed to be quite reliable for measuring different aspects of doctor's communication with real patients.

6.1.10 Validity of the CARE measure for students

Valid assessment of clinical competence aims to test what the doctor (or the senior medical student) actually does in the workplace. The CARE measure is a real-life clinical assessment method, which aims to assess performance at Miller's 'does' level and predicts day-to-day performance in clinical practice well. The actual clinical practice settings ('*in vivo*'; (Epstein and Hundert, 2002) reflects the real behaviour of the student, who may behave differently in examination settings such as the OSCE, where standardised patients are used ('*in vitro*'; (Boulet et al., 1998). Thus, the CARE measure has encouraging face validity.

Estimates of validity are dependent upon the nature of the population sample being measured and, to lesser or greater degree, the circumstances under which they are being assessed. Therefore, a comprehensive literature review on the CARE measure was conducted to provide evidence of validity for any measure that was built over time, as validations occurred in a variety of populations (Kimberlin and Winterstein, 2008). This is also was applied on the CARE questionnaire when its validity was tested in different populations (GPs, consultants, anaesthetists), settings (primary care centres, secondary-care centres) and languages (English, Chinese, German). Thus, the CARE measure has now been shown to have adequate construct validity in the postgraduate context.

The construct validity of the CARE measure had been tested thoroughly before, and it was found that this tool has robust internal structure (Mercer and Murphy, 2008). The details about the curricular validity of the CARE measure were demonstrated and discussed in the mapping exercise chapter. The CARE measure's concurrent validity was tested by correlation with other measures that have been previously validated. The scores obtained from the CARE measure were directly related to scores obtained from

more established measures of the same variable (different aspects of consultation skills). The findings of this study showed a moderate but statistically significant correlation of the CARE measure with GP OSCE and tutors means at the end of the GP block (these measures sharing presumably related constructs). Thus, we have to acknowledge that either CARE is less good than the faculty-based measures or perhaps measures something different, for instance a discrepancy between what patients and doctors value in a students' performance. However, the high number of patients who participated in this study and their willingness to rate student's relational empathy reflected the acceptability of this measure across the socio-economic spectrum and is another important attribute for any tool such as this.

6.1.11 Reliability and educational impact

The relative importance of reliability versus educational impact depends on the purpose of the assessment. As the CARE measure focuses largely on providing the trainee/student with feedback to inform their personal development, planning should focus on educational impact, with less of an emphasis on reliability. In contrast, a high-stakes examination needs high reliability and validity at the expense of educational impact. Assessment tools at the 'shows how' level of Miller's pyramid such as OSCE are undertaken outside the real clinical environment. On the other hand, assessing the 'does' level of Miller's pyramid is considered ideal even if it is challenging to do so during consultations with patient in primary or secondary settings. The CARE measure offers a solution and is the only available real patient outcome measure that offers a feasible means of measuring the 'does' element. However, it must be recognised that even this may not reflect actual practice when the subject is not knowingly being assessed, which would require covert observation.

6.1.12 Patient, student, and staff feedback

The face and content validity of the CARE measure was supported by the patients' feedback from comments on CARE questionnaire. In addition, it is supported by the feedback from the expert group of teaching staff who participated in the mapping exercise, who scored other assessment tools in addition to the CARE measure during the clinical phase of training at Dundee Medical School. Another source of support for the face validity of CARE was the students themselves, through contact with them throughout the period of GP block and, in particular, the feedback meetings which were conducted at the end of GP block.

It is important to offer ways of improving performance to those students with below average scores at the time of the assessment. This can be achieved by sending the feedback to students and by doing so within a reasonable period of time before the end of the block. In addition, more training on weak points of consultation skills or empathic skills has generally been found to be of benefit (Griffin et al., 2004).

The findings of the study revealed that the percentage of negative comments was low. Negative comments provide a very important kind of feedback, since both positive and negative feedback can be used to enhance learning (Kluger and DeNisi, 1996), although some caution must be taken against the overuse of negative comments as they may threaten learners' self-esteem and self-efficacy (Hattie and Timperley, 2007); (Shute, 2008). In this study, the ratio of positive to negative feedback interventions and overall specific feedback comments (positive and negative) may be considered acceptable in terms of being effective feedback. However, we suggest more research should be done in this area because of its importance for student learning.

Sargeant and colleagues (2005) found that using scores alone to provide feedback was of less value in changing behaviour than when combined with free text as a kind of multisource feedback. Therefore, a comment box also was included with the CARE questionnaire for free-text comments at the end of the form by Mercer and his colleagues (Mercer et al., 2008b). In that study, the combination of patients' scores and free text provided the anaesthetists with feedback to assess their development needs and performance. The percentage of patients' comments in the present study was higher than the percentage of comments collected by Mercer and his colleagues (Mercer et al., 2008b), which reflected an apparent desire on the part of patients to develop these students, whether with positive or negative comments. This is perhaps not surprising, as it might be imagined that patients willing to consult with students would wish to do what they could to contribute to the resultant learning for that student.

The comment box added another view of the patients towards clinical consultations with students and provided educators with valuable feedback on the clinical training process. Also, at the same time, it gave feedback to the students themselves about the strength and weakness areas of consultation skills from the patients with whom they conducted independent consultation. This contributes to the recent guidance from the GMC encouraging sharing the assessment process with patients and giving them an opportunity to provide educators with constructive feedback to help identify both strong and weak areas in students and in the curriculum (GMC, 2009a).

The overall reliability of the CARE measure for students gives educators confidence in its ability to provide scores as a basis for the development of a system of feedback to students on their empathy when consulting. However, the provision of scores alone has been shown in other contexts not to lead to behaviour changes and is likely to be

rejected (Sargeant et al., 2005). Here in the present study, when provided with a combination of measure scores and free text when available, the CARE measure could offer feedback to help identify student's development needs in this area of their performance. This kind of feedback (scores and free text) may provide students with insight and promote change where applicable (Sargeant et al., 2005).

6.2 Strengths and limitations of this study

6.2.1 *Strengths*

An important strength of the present study was that we attained a high response rate among students in the second and third years of the study period, although it was in some ways disappointing during the first academic year. In addition, the majority of the participating students were able to collect the required number of CARE questionnaires. The number of students who took part was sufficient to detect differences in CARE measure scores between students with an acceptable degree of reliability in this case. The high number of patients who rated the participated students in this study might also be considered as a strength, and it is close to the number of patients that participated in the Mercer study in general practice (Mercer et al., 2005b). The present study provides performance data on the CARE measure in a large sample of student general practice consultations in the primary care setting for the first time.

The students' feedback was helpful and important for developing the CARE questionnaire in teaching and training in the future. Another key strength which was not available in postgraduate studies was the comparison of the CARE measure with the existing undergraduate assessment tools (e.g. GP OSCE, Mini-Cex, etc.). In summary, this study showed a new use of the CARE measure with medical students with adequate data to evaluate its reliability and validity as well as highlight some of the issues to be considered regarding its introduction and educational utility.

6.2.2 *Limitations*

The following points are recognised weaknesses of the study. A potential weakness of this work is the unbalanced student sample. Only those who chose to take two- and three-month GP attachments were eligible, and most of them chose a longer training

period because they were interested in their career. Thus, they may be atypical in their willingness and also in scores they achieved.

Another of the limitations of the study includes the fact that patients were not recruited entirely at random. Patients were excluded by the practice staff and tutors according to the instructions provided. This might exclude examination of some consultation abilities (i.e. missing consultations with complicated medical cases), and when students or tutors were involved in distributing the questionnaires, they may have biased the results by omitting patients they suspected might give low scores (see below). The students' contextual data (age, gender, ethnicity, etc.) was not collected and could be a useful focus for future studies.

Not all of the CARE forms were distributed to patients as planned (by a reception staff member); in three of the 66 students, either the forms were distributed by the GP tutors or the student himself/herself recruited patients. This alternative procedure of questionnaire form distribution by the tutor or the student was followed only when the main method of distribution was not possible and, depending on the reasons for this, might be a source of data bias; this study was originally designed so as not to let the student know which patient he/she rating him/her as this could have affected the student's behaviour (e.g. if the student knew that a particular patient would be evaluating him/her, the student might behave in a different way). But this could not be verified or measured and we did not have access to personal and demographic details of the consulting patients; thus we cannot say whether those who completed the questionnaire were representative of attendees in general.

In addition, the present study did not collect information on how many patients refused to participate (patient response rate); thus, the failure of some students to collect the

required number of forms for this study might related to a low throughput of patients, a high refusal rate, or simply a lack of motivation within certain clinics to hand out the questionnaire. However, the overall response of patients was adequate, and, as in previous studies on the CARE measure, there were no major problems recruiting patients that were reported to the investigator (Mercer et al., 2008b, Mercer and Murphy, 2008).

6.2.3 Summary

In summary, this study has demonstrated that it is feasible to utilise the CARE measure in undergraduate medical education with final-year students in the GP context. The psychometric properties of the tool are consistent with previous performance, although the patients appear to consistently score students higher than they do their doctors. Student scores correlate with relevant related assessments offering important concurrent validity which has not been presented before. In addition it has been shown that this patient-based measure can identify students with weak consultation skills and also identify the reasons for those weaknesses based on an analysis of the areas in which their scores are low as well as examining patient free-text comments. Hence, although the CARE measure has been shown to perform well enough for summative use, its main potential may be in the educational impact on those requiring assistance to develop their skills.

7 Mapping exercise

7.1 Background of mapping exercises

For the purpose of this thesis, a mapping exercise is defined as the investigation of the capacity of a suite of different tools of assessment to cover the scope of a defined set of desired educational curricular outcomes. As such, a mapping exercise is a test of the validity of a curriculum's assessment system to cover the curriculum's specified educational objectives. The methods used in this thesis are based and replicate previous research exercises undertaken to inform the development of workplace-based assessment for UK postgraduate general practice education (Murphy et al., 2008) and a recent study to define a suite of tools used to research the professionalism of established UK general practitioners in reacting to performance feedback (Murphy et al., 2012).

For the purpose of this thesis, the mapping exercise tested the perceptions and agreement of both faculty staff and students on how well the existing suite of available 2011–2012 academic year student assessments with the inclusion of the CARE measure were thought to cover the testing of the GMC's required student educational outcomes as specified in the GMC document *Tomorrows' Doctors* (GMC, 2009a). *Tomorrows' Doctors* attributes are subdivided into three domains: the doctor as 1) a scholar and a scientist, 2) a practitioner, and 3) a professional.

7.2 Introduction and rationale

The attitudes towards medical schools' educational programmes have changed since the 1990s. At that time, the newly developed outcome-based education was adopted by medical schools in the UK (Harden, 2002). Medical schools altered their educational

programmes to be in line with these developments. More focus was given to improving the quality of health care by systems that reviewed specified learning outcomes. All the medical schools in Scotland, including Dundee Medical School, adopted the ‘learning outcomes’ approach in the early twenty-first century (Simpson et al., 2002). This led to shift in medical schools’ curricula from traditional approaches of education (knowledge and recall) towards learning outcomes such as problem solving, clinical judgment, communication skills, and professionalism (Shumway and Harden, 2003).

Currently, however, medical school curricula do not include patients’ feedback into the suite of assessments made available to medical trainees (future doctors) to provide feedback for their quality improvement. The challenge for current curricular assessment systems is whether they offer a broad enough suite of available feedback to meet today’s students’ needs and the specified requirements of training as set by GMC attributes in *Tomorrow’s Doctors*. *Tomorrow’s Doctors* attributes fall into three main categories (GMC, 2009a) as follows:

- five attributes under ‘The doctor as a scholar and a scientist’
- seven attributes under ‘The doctor as practitioner’
- four attributes under ‘The doctor as a professional’

Graduates of medical schools in the UK must achieve all the above 16 attributes. Medical schools need to set their assessment process and clarify the learning outcomes and to find the most appropriate tool or tools to assess each attribute. The planning and implementation of an effective assessment process can only be achieved if there is a clear definition of required learning outcomes (Shumway and Harden, 2003), and the GMC *Tomorrow’s Doctors* document provides this template for curricular

development. As a result, in the UK, medical schools develop and implement their comprehensive assessment strategy to meet the updated *Tomorrow's Doctors* requirements. The award of a medical degree by a UK medical school means that the graduate has been appropriately trained and is fit to practice medicine to the standards set by the GMC. Accordingly, medical school graduates in the UK should meet the requirements for their future careers as physicians. The GMC have specified patient completed consultation satisfaction questionnaires as a key required component for the revalidation of doctors in the UK (GMC, 2002). The CARE measure has been accepted as an instrument for this purpose.

To ensure the ongoing quality of the medical schools' assessment systems, it is essential for medical schools consider the inclusion of dependable and trustworthy patient feedback assessment programmes incorporated into medical education. There is no single measuring tool that can adequately assess the doctor's clinical performance and, as a result, there is increasing focus on the value of a suite of assessment tools to provide a system to provide a more complete picture (Van der Vleuten and Schuwirth, 2005). As a result, the validity and reliability of the assessment results are enhanced by using multiple instruments (Van der Vleuten, 1996). In the last decade, views on the purpose of assessment and the methods used were transformed by this consideration of the assessment process as whole (Shumway and Harden, 2003); (The Scottish Doctor, 2007). This was an important step towards delivering an outcome-based education, where a student's competency is tested against each individual learning outcome (GMC, 2009a).

7.3 Aims

The aim of this exercise is to identify to what extent different medical student assessment tools cover the GMC 2009 *Tomorrow's Doctors* attributes for graduates. The ultimate goal of this mapping exercise is to evaluate the curricular validity of the CARE measure and assess if it adds value and/or complements existing students' assessments.

7.4 Methods

7.4.1 *Study design and ethical issues*

This study's data collection exercise was facilitated by the use of a web-based questionnaire designed to collect the perceptions of medical teaching staff about a suite of currently used assessment tools that include the CARE measure. There was a query if this exercise required ethical approval because it has different population sample (teaching staff) from those in the previous study of the CARE measure validation in earlier chapters of this thesis. Therefore, we consulted Dundee's University Research Ethics Committee for advice, and they agreed that, as the proposed study was an audit of teaching, it did not require approval.

7.4.2 *Questionnaire used in the mapping exercise*

Participants were asked to complete a seven-point Likert rating scale (graded from 1 (poor) to 7 (excellent)) to record their perceptions of the utility of each assessment tool against the GMC specified training attributes. There were 11 different assessment tools listed as columns in the exercise (Appendix 12). These were as follows:

- fourth-year Extended Matching Item (EMI) format test designed to assess the clinical application of knowledge base

- fourth-year OSCE: this is a 12-station clinical OSCE that includes consultations, examinations, procedures, and data interpretation
- fifth-year case report: this focuses on seven attributes (e.g. ethics, patient management); case-based essays are marked for inclusion in students' curricular portfolios
- fifth-year block feedback form: this is a Scottish Doctor 12-outcome-based form completed by head tutors at the end of each clinical attachment
- final-year OSCE: this is a seven-station (15-minute each) clinical OSCE oriented towards assessing foundation/year one-procedures, consulting, prescribing, etc.
- Portfolio exam: this is a final-year assessment of a collected, written portfolio, including grades for all assessed work, clinical attachments, etc., plus a 40 minute viva
- Progress test: this is an online, knowledge-based test requiring free-text responses to around 250 items.
- CARE questionnaire: this 11-item, patient-completed form was collected following student-led consultations in general practice (for more details see Section 4.3)
- mini-CEX: this is a structured observation and feedback form used to guide evaluation of student's consultation and/or examination skills (for more details see Section 4.7)
- case-based discussion: this is a structured feedback form used to guide assessment of discussion regarding a patient seen in general practice

- Direct Observation of Procedural Skills: this is a structured observation and feedback form used to guide evaluation of an observed procedure, such as intravenous cannulation

The electronic mapping exercise questionnaire rows listed 16 specific GMC attributes which acted as descriptors for three GMC domains (or sections) of interest.

The first section asked about ‘the doctor as a scholar and a scientist’, using five related questions, including questions addressing the application of medical practice biomedical scientific principles, methods, and knowledge relating to different basic sciences (e.g. anatomy, biology, bacteriology, etc.); the application of psychological and social sciences principles, methods, and knowledge to medical practice, population health, and the improvement of health and health care; and the application of the scientific method and approaches to medical research. The second section requested information on the ‘doctor as a practitioner’ by asking seven questions related mainly to carrying out consultations with patients, diagnosing and managing clinical presentation, communicating effectively with patients and colleagues, providing immediate care in medical emergencies, prescribing drugs and carrying out practical procedures safely, and using information effectively in a medical context.

The third section dealt with ‘the doctor as a professional’. In this section the graduate was assessed on the ability to behave according to ethical and legal principles, reflect, learn, and teach others. In addition, the graduate was assessed on the ability to learn and work effectively within a multi-professional team, to protect patients, and to improve care (Appendix 12). The electronic form of this exercise had a ‘hover button’ for the tools at the top. This meant that if anyone moved their mouse over the title of

the tool, the descriptor for the assessment tool would show; and if they clicked on the title of the tool, a copy of the descriptor would appear.

7.4.3 Piloting the mapping exercise

There was concern about some practical points in completing the electronic form of the mapping exercise and the time needed from each participant to fill the scores precisely and in reasonable period of time. Therefore, a pilot was undertaken prior to the process of data collection. This was conducted by distributing the mapping exercise questionnaire to two staff members from the Dundee medical undergraduate department. They were asked to complete the exercise and comment on the practicality after filling in the scores. Both encountered no technical problems but suggested an important point regarding the tools templates:

‘I found this an interesting exercise. Data entry was straightforward. I would have found it easier if I had studied the templates for each tool prior to responding. Perhaps this should be recommended more strongly.’

It was important at the pilot stage to get feedback on practicality of the exercise. Both of the teaching staff who participated in the pilot found that the exercise could be completed within 10–15 minutes, which they found feasible. They suggested that one could complete the exercise in stages by intermittently saving their partially completed which could then be finished at a later time. The staff also expressed concern that different people might have different levels of awareness of the various assessments, and that this might influence their scores.

7.4.4 Exercise sample and data collection

The data collection process of this mapping exercise was carried out between October and December 2011. The exercise was completed on behalf of the faculty of medicine and was supported personally by the Teaching Dean of Dundee Medical School, who agreed to distribute the exercise in his name. As a result, the exercise became a faculty issue, which caused it to attract more interest from the invited participants. The exercise recruited teaching staff members from the University of Dundee Medical School. Generally, those engaged directly with assessments and had experience with the existed exams and were specifically familiar with instruments. Accordingly, a list of 35 names from Dundee medical faculty staff (block leaders and undergraduate medical educators) was prepared for this exercise. The names and emails addresses were obtained from the Dundee medical secretary office. The principle investigator sent an email on behalf of the Teaching Dean of Dundee Medical School to each one on the list, inviting him/her to take part in this exercise, asking them to give their opinion on how well we currently assess the GMC 2009 *Tomorrow's Doctors* attributes across the clinical years, and specifically to inform them about the patient-based measure (the CARE questionnaire). A reminder email was also sent to those who did not reply within two weeks, encouraging them to complete the exercise as their opinions were considered to be of great importance for future assessment design. The email included a description of the purpose of the exercise and a web link (see Appendix 11). In addition, each participant was provided with a specific username and password to log in his personal form. After the participants had logged into the exercise main page, another guideline paragraph on the top introduced the technical instructions, to help the participants to find information on any tool through a fast 'help' button and to save time by going through each tool horizontally. At the end of

the exercise, the participant was prompted to submit his/her score. A note was added to the cover page of the exercise asking participants to fill in their scores vertically down each column of the presented exercise. This ensured that participants were familiar with each assessment tool in turn and could focus on training their perceptions of its coverage of the presented GMC attributes (see Appendix 13).

7.4.5 *Statistical analysis of the mapping exercise*

Data from the completed exercises was collected, and an Excel report with the raw data was prepared by members of the IT staff and was then made available for analysis. G-theory was used to investigate the level of agreement between participants on how well tools were perceived to test outcomes (Streiner and Norman, 2008). The sources of variance in the mapping exercise of the present study were investigated via a D-study. ‘Tools’ (subjects of the exercise) was the facet of differentiation, and ‘participated staff members’ (raters) and ‘attributes’ (GMC attributes) were the facets of generalisation. The one-way ANOVA test was used to analyse the mapping exercise data to determine any statistically significant differences between the students’ assessments tools or groups of these tools in their respective capacity to test each of the 16 GMC *Tomorrows’ Doctors* outcomes. The Tukey post hoc test; which was provided by the SPSS software package, was used to investigate how well each tool (column) tested each GMC attribute (row; i.e. which was the best or least statically significant tool for measuring each GMC attribute (see Table 7.1)).

7.5 Results of the mapping exercise

7.5.1 *Response rate and level of agreement*

The total number of those who completed the exercise and sent their scores back for analysis within the two months allocated for the exercise was 17. These included nine members of the Dundee faculty staff block leaders and eight from the undergraduate teaching staff. Those who declined to participate did so either due to personal reasons or due to unfamiliarity with the presented assessment tools. The level of agreement among participants was high ($G = 0.7$). This was reassuring but unsurprising, given the high number of raters used ($n = 17$), and gave credibility to the study's results by ensuring that mean scores obtained pointed to conclusions drawn from the exercise offered a reassuring level of content validity.

7.5.2 *Coverage of GMC attributes by assessment tools*

The coverage scores of each GMC attributes by the 11 chosen assessment tools for this exercise are shown in Table 7.1. Bold green scores denote the highest and statistically most significantly ($p = 0.05$) valued tool or group of tools for the attribute considered, and red italics denote the least and statistically significantly valued tool or group of tools. For example, according to the rating of the experts who participated in this exercise, the attributes of the 'doctor as a scholar and scientist' category showed that the fifth-year case report had higher significant means than all other assessment tools. The fourth-year EMI was one of the most valued tools for measuring the first GMC attribute. The CARE was the least valued tool for assessing most of the first of the 16 GMC attributes.

For the second domain of the GMC attributes, ‘the doctor as practitioner’, the fifth-year OSCE showed remarkable perceived importance as the highest significant measure for the entire group. In contrast, the progress test and fourth-year EMI were the least valued for most of the attributes of this domain. In addition, some of the GMC attributes were significantly valued by more than three tools. For example, eight tools, including the CARE measure, were found to measure the attribute ‘communication with patients’ with high significance. In addition, the attribute of ‘use of information effectively’ was found to be significantly valued by six tools. In contrast, the attribute ‘provide immediate care in medical emergencies’ has the greatest number (9) of least significant tools. Only fourth- and fifth-year OSCEs measured this attribute significantly. The third category of GMC attributes, ‘the doctor as professional’, had the lowest number of valued measuring tools for each attribute, except for ‘behaving according to ethical and legal principles’, which was significantly valued for seven tools including the CARE measure. In contrast, the progress test was the least valued tool for testing the attributes of the third category.

The results of this exercise also indicated the inability of tools to measure some GMC attributes, as only one significant instrument was available for attributes such as ‘apply scientific method and approach to medical research’, ‘reflect, learn and teach others’ and ‘learn and work effectively within a multi-professional team’. The perceived relative over-testing of some attributes by multiple tools and the under-testing of other GMC needed attributes was an important finding for the current system of curricular assessment and is discussed later.

7.5.3 *The CARE measure in the Mapping Exercise*

The mapping exercise results were arguably intuitive. The CARE measure would seem to have little to do with the attributes of ‘the doctor as a scholar and scientist’, as it deals with the consultation process and not with the knowledge of basic scientific information. The CARE measure showed a relatively higher means in assessment of the GMC attributes of ‘doctor as a practitioner and professional’, more specifically, in measuring ‘carrying out consultation with patient’ (5.18) and ‘communicate effectively with them’ (5.65). Two other assessments yielded similar scores in these areas (fourth- and fifth-year OSCE). In addition, ‘the behaviour of trainees’ (4.00), and ‘protect patient and improve care’ (3.47) exhibited more neutral scores on the tested (1–7) scale for the CARE measure. Participants thought that CARE tested the following four GMC headings well:

- 1) carry out a consultation with a patient
- 2) communicate effectively with patients and colleagues in a medical context
- 3) behave according to ethical and legal principles
- 4) protect patient and improve care

Participants thought that CARE tested the following four GMC headings poorly:

- 1) apply to medical practice biomedical scientific principles, method and knowledge relating to: anatomy, biochemistry, etc.
- 2) apply to medical practice the principles, method and knowledge of population health and the improvement of health and health care.

- 3) provide immediate care in medical emergencies
- 4) prescribe drugs safely, effectively, and economically

Table 7.1. The CARE mapping exercise- Phase 3 assessment tools versus GMC attributes (n=17).

GMC Tomorrow's Doctors Outcomes 2009	Phase three assessment tools											
	Year 4 EMI	Year 4 OSCE	Year 5 Case Report	Year 5 Block Feedback	Year 5 OSCE	Portfolio Exam	Progress Test	CARE	Mini- CEX	CBD	DOPS	Mean Scores
1. The doctor as a scholar and a scientist. The graduate will be able to:												
1. Apply to medical practice biomedical scientific principles, method and knowledge relating to: anatomy, biochemistry, etc.	4.82	3.12	4.71	2.12	2.94	3.76	4.65	1.47	3.18	4.35	2.29	3.40
2. Apply psychological principles, method and knowledge to medical practice.	3.59	3.65	4.82	2.53	3.35	3.94	2.76	2.88	3.24	4.47	2.18	3.40
3. Apply social science principles, method and knowledge to medical practice.	3.53	3.18	4.65	2.41	2.82	3.88	2.47	2.47	2.82	4.29	1.82	3.12
4. Apply to medical practice the principles, method and knowledge of population health and the improvement of health and health care.	3.59	3.12	4.53	2.47	2.53	3.47	2.94	1.65	2.47	3.88	1.71	2.94
5. Apply scientific method and approach to medical research.	2.59	1.35	4.00	1.82	1.71	3.29	1.94	1.24	1.65	3.00	1.12	2.16
2. The Doctor as a practitioner. The graduate will be able to:												
1 Carry out a consultation with a patient:	1.53	5.82	2.24	3.47	6.29	2.88	1.18	5.18	5.35	4.35	2.82	3.74
2. Diagnose and manage clinical presentations.	3.29	5.35	3.88	3.65	5.76	3.71	2.47	3.18	5.12	5.24	2.88	4.05
3. Communicate effectively with patients and colleagues in a medical context.	1.41	5.59	3.35	4.59	6.06	4.41	1.47	5.65	4.76	4.53	4.00	4.17
4. Provide immediate care in medical emergencies.	1.82	5.35	2.41	2.59	5.41	2.88	1.88	2.29	3.29	3.29	3.06	3.12
5. Prescribe drugs safely, effectively and economically.	3.12	4.29	3.00	2.71	5.24	3.24	2.35	2.06	3.88	4.24	2.53	3.33
6. Carry out practical procedures safely and effectively.	1.35	4.76	1.76	3.06	5.12	2.59	1.29	2.53	2.65	2.53	6.00	3.06
7. Use information effectively in a medical context.	3.06	4.76	4.47	3.41	4.76	4.24	2.06	3.00	4.29	4.88	2.65	3.78

3. The doctor as a professional. The graduate will be able to:												
1. Behave according to ethical and legal principles.	2.29	4.53	3.35	4.47	4.76	4.47	<i>1.82</i>	4.00	4.06	4.24	3.59	3.78
2. Reflect, learn and teach others.	<i>1.35</i>	2.12	3.65	3.12	2.24	4.76	1.65	2.94	3.18	3.88	2.82	2.88
3. Learn and work effectively within a multi-professional team.	1.59	2.53	2.59	4.29	3.00	2.82	<i>1.35</i>	2.47	2.82	3.29	2.88	2.70
4. Protect patients and improve care.	1.88	3.12	3.12	3.35	<i>3.65</i>	4.00	<i>1.76</i>	<i>3.47</i>	<i>3.65</i>	4.00	<i>3.82</i>	3.26
Mean scores	2.55	3.92	3.53	3.13	4.10	3.65	2.13	2.91	3.53	4.03	2.89	3.31

*Tools or group of tools significantly different from the rest as being the **most highly valued** for each attribute are represented in **bold** green font.

*Tools or group of tools significantly different from the rest as being the *least highly valued* for each attribute are represented in *italic* red font.
(p=0.05)

7.6 Discussion and conclusion of the mapping exercise

The exercise was introduced to the teaching staff and members of the undergraduate learning centre at Dundee Medical School, who were familiar with all the assessment methods. They were all familiar with the assessment tools presented and had been briefed on the content of the additional assessment (the CARE measure). The high participant agreement gave credence to the results obtained, and robust statistical methods were used to analyse the study data. This high level of agreement among participants in this exercise was not surprising, given the corresponding background of participants concerning the assessment tools and also the significant number of raters involved ($n = 17$). It is important to appreciate that while the inclusion of the CARE measure was new to the current curricular system of assessment, the CARE measure is widely used in the postgraduate sphere throughout the NHS and is used and approved for medical revalidation in the UK (Mercer, 2009). In addition, some of the participants had completed the questionnaire as part of a previous study on workplace-based assessment in postgraduate medical education and training in the UK (Murphy et al., 2009b).

7.6.1 General review of Dundee Medical School assessment programme

The chapter containing the literature review for this thesis (Chapter 2) highlighted the importance of the utility of the assessment programme as a whole. No single method can be expected to assess all the attributes required of a doctor, and there is a need for the combination of a well-chosen suite of assessment instruments to gain a complete picture on performance and competence of a group of candidates (Van der Vleuten and Schuwirth, 2005). It was noticed from the results of this mapping exercise that most of the GMC attributes were assessed by two or more of the Dundee Medical School

instruments. This was reassuring in its coverage and demand on resources. However, other attributes were only assessed significantly by one tool (e.g., the fourth and fifth GMC attributes, which are concerned with application of principles, method, and knowledge for improving healthcare and the application of scientific methods and approaches to medical research, respectively, as well as the attribute under the category of ‘the doctor as a professional’ concerning learning and working effectively within a multi-professional team). These attributes specifically should be addressed by educators when re-evaluating the assessment programme to try to fill this gap or weakness in the system. On the other hand, other assessments that were least highly valued (e.g., the progress test) might need to be reconsidered as to their usefulness within the whole assessment programme.

Key lessons for the current system of curricular assessment can be learned from the results of this mapping exercise. The findings offered Dundee Medical School an opportunity to refine efforts to best utilise resources and achieve a broad coverage of needed GMC attributes by perhaps a minimum of two different methods. Arguably, communication skills are currently over tested, and a reduction in time and resource in this area could allow more focus on the development of better testing in less well-covered areas such as health improvement and scientific methodology.

7.6.2 The role of the CARE measure within the assessment programme

The CARE measure offers a curricular assessment system a ‘patient’s voice’. This cannot be captured by other surrogate methods. Its inclusion would bring an authenticity currently lacking in the assessment programme and fits with the future postgraduate expectations for medical revalidation. In this study, CARE was one of the

most highly statistically significantly valued tools for some of the most important GMC attributes as well as for the four GMC domains, as it gave patient support to a suite of testing by more than one tool. Schuwirth and Van der Vleuten, in their study on changing education, assessment, and research, advocated the role of assessment programmes rather than using individual measuring tools and supported the idea that no one tool could be relied upon (Schuwirth and Van der Vleuten, 2004b). In addition, Murphy and his colleagues demonstrated that the CARE measure could cover all of the qualities required of a doctor when combined with other competence-based assessments (Murphy et al., 2008).

The limitations of the CARE measure in measuring some of the GMC attributes were manifested as low scores provided by participants who believed that this tool had nothing to do with those attributes in particular. The weakness of the CARE questionnaire in measuring the ability of a doctor to provide immediate care in emergencies and prescribing drugs safely was understandable and not of concern, as there are other tools in the system of assessment that measure these attributes (fourth- and fifth-year OSCE). Therefore, as concluded from this mapping exercise, the CARE measure can add unique value to the validity of the medical school's curriculum, and based on this thesis's findings, I recommend its inclusion into the curriculum to help steer doctors that are ready to meet their future career challenges and expectations of their patients.

8 Implications and Conclusions

8.1 Implications

The implications of this thesis are significant for practice, medical education, student learning, and future research. Underpinning these implications are the thesis's findings that an authentic measure of patient empathy, as assessed by patients, offers reliable and feasible feedback on which students can base their appropriate reaction, learning, and change for better professional practice. In addition to helping to set students out on an effective and professional path in their practice, it introduces to them the importance of patients' voice in helping them maintain their professionalism career long.

8.1.1 Implications for practice and medical education

The influence of examinations on students' learning has been increasingly emphasised by test developers. The concept of assessment driving learning requires consideration during an assessment tool's development. This consideration can offers opportunities to align assessment to reality to best prepare students for their professional careers. Here, the CARE measure has the advantage of testing and giving students educational feedback on their empathic skills with authentic. In addition to driving learning by providing mean scores on arrange of empathetic behaviours and skills, arguably CARE drives learning by providing students with a narrative form of patient feedback through their attached comments. This feedback is a unique insight for students that cannot be as authentically captured by any surrogate or simulation. In doing the above, the CARE measure assesses students at the apex of Miller's pyramid, the 'does', and so meets the international challenge for tools involved in testing clinical competence of students' performance (Wass et al., 2001). Also, assessing a student at the highest level

of Miller's pyramid requires an evaluation of the student's habitual performance in everyday practice, and this matter was accomplished by utilising the CARE measure.

The consequences of the assessment tool are important as this determines the impact of scores on learners. The CARE measure as a high-stakes assessment tool might lead the students to think that if they did poorly they would fail the whole year and therefore suffer from performance anxiety during consultation. Conversely, student might think that the CARE measure scores were unimportant if it were used as a low-stakes assessment.

In practice, the different aspects of utility (reliability, validity, educational impact, acceptability, cost, and feasibility) need readjustment of their weight for each individual situation in regards to the context and purpose of the assessment (Van der Vleuten, 1996). If the CARE measure was used as a high-stake assessment, as in the case of summative assessment, reliability should be given have higher weight than other variables in the utility equation. On the other hand, if the CARE measure were used only on a formative basis to help train medical students, the threshold of needed reliability (and number of patient questionnaires needed) would be reduced. As it is acceptable that the reliability of the instrument can be lower in case of formative than for summative assessment, the CARE measure can be used in a formative assessment form with fewer questionnaires per each student, i.e. 20 patients' completed questionnaires per student are needed for reliability 0.7 according to the results of the present study. On the other hand, if medical educators wish to apply the CARE measure for high-stakes assessment, the adequate level of reliability should be above 0.8, and at that time the number of questionnaire forms per students must be 35 forms or more.

In their systematic review, Hemmerdinger and colleagues (2007) concluded that CARE was the only measure that had adequate validation research, including criterion-related validity with other patient-centred measures of quality of care. They stated that the CARE measure is the only patient-rating (second-person) instrument which has evidence of reliability, excellent internal consistency, and showed comprehensively validation for postgraduate doctors (Mercer et al., 2004a). The present study added more evidence of reliability and validity of the CARE measure for medical students in addition to some other educational characteristics such as feedback and as a tool for self-audit (e.g. in appraisal and revalidation).

8.1.2 Implications of patient feedback on student learning

The qualitative components of many surveys are usually ignored and no one is interested in reading long lists of narrative answers to open-ended questions (Chi, 1997). In contrast, in this study, we were aware of the importance of our qualitative data as it comes from real patients and reflects their perspectives on the future doctors and the quality of health care services. Thus, the patients' comments were given high attention.

The feedback information that the CARE measure provided us with students' performance during clinical consultation was of great importance in promoting positive empathetic skills and development of desirable behaviour in students. However, this effect of feedback is complex and sometimes may have resulted in an outcome that was opposite to what we intended to promote and develop (Kluger and Van Dijk, 2010). Although the comments of the patients were largely positive and

non-specific, it is accepted that such narrative feedback is still effective to enhance learning in our high achievers senior medical students (Moreno, 2004).

Recently Boud and Molloy's thoughts about feedback have changed many previous assumptions, such as the idea that all feedback is good or that more feedback means better results, which mask the complexity of feedback (Boud and Molloy, 2012, Boud and Molloy, 2013). In addition, the last decade showed marked increase in studies that call for using patients' voice in medical education (Towle, 2006); (Howe, 2007); (Jha et al., 2009). In their study, Reinders and colleagues (Reinders et al., 2008) reported that patients' feedback on consultation skills was useless as it contain biased information. In addition, there is consensus that students in higher education are dissatisfied with such feedback (Boud and Molloy, 2012); (Ferguson, 2011). However, researchers and educators are still convinced about the potential benefits of such feedback and feel that it can be used to encourage students' learning (Nicol, 2011); (Boud and Falchikov, 2006). In addition, learners can use the feedback as an interactive process that provides them with a clear view of their performance (Clynes and Rafferty, 2008). In this context, it is important to acknowledge that students who participated in this study showed high interest in patients' comments as evidenced by the fact that some of them attached CARE scores and comments to their portfolio.

The work reported in this thesis was conducted in clinical educational settings where error avoidance is very important for patient's safety and optimising patient's health. Patients as one of the feedback providers have an important role in enhancing and developing the professional skills of physicians and medical students in practice. It is accepted that formal educator-led feedback may be more effective when combined with the informal feedback delivered by patients (Boud and Molloy, 2012), and recent efforts by educational researchers to formalise the patients' role in enhancing

students' learning, especially of consultation skills, has been highlighted (Lown et al., 2008); (Reinders et al., 2008, Reinders et al., 2010). Crucially, patients' comments, as made available by this thesis work, offer students the opportunity to develop empathic skills by taking an active role to improve, where appropriate, any needed areas to meet patients' needs.

8.1.3 Implications for the Dundee Medical School assessment programme

The mapping exercise provided a good indication on the utility of the CARE measure as one of the assessments performed in the final year in Dundee Medical School. The results of the mapping exercise were crucial in supporting the adoption of CARE by the medical school. The findings of the mapping exercise provided Dundee Medical School with interesting information and an opportunity to review and modify the assessment programme to comply with GMC attributed in order to meet the updated *Tomorrow's Doctors* requirements. In addition, the exercise had an important role in specifying areas of weakness in the assessment programme and bringing them to the educators' consideration. Thus, the educators in Dundee Medical School have an opportunity now to refine their efforts for best utilisation of resources and achieve a broad coverage of needed GMC attributes.

8.1.4 Implications for future research

Further qualitative work is required on the use of the CARE measure to assess students in other settings, such as secondary-care institutions, to see if the pattern is the same or different in different settings. There is also a need to study the reliability and validity of the CARE measure with simulated patients (as standardised patients) in clinical

encounters to identify the level of students' empathy and demonstrate any changes in empathic level for students as they go through medical school. In addition, it will be necessary to learn the effects of patients' feedback on students' performances in clinical consultations. For example, it would be useful to know how much positive comments encouraged students, or what the effect of critical comments on the students was. Consequently, the impact of patients' feedback on students' performance in primary settings needs to be investigated thoroughly in the future. Applying the CARE measure to medical students consulting hospitalised patients was useful, but, the process of distribution and collecting the questionnaire forms might be too complicated and difficult to run smoothly in hospitals. This could be an interesting challenge for future research on the CARE measure.

8.2 Conclusions

This thesis has identified, applied and evaluated a patient completed measure of senior undergraduate medical students' consultation skills. The CARE measure has been shown to be reliable, valid, and feasible in this context within the Dundee curriculum. However the way in which CARE might be introduced or best used on a routine basis requires careful thought. There is considerable overlap with existing elements of assessment, the OSCE in particular and a considerable logistical effort is required to gather and process the scores, and this would be exaggerated should it be used in a summative capacity where students may have reason to select patients or even submit falsified responses. So, although there is potential to use CARE as part of an examination system, it appears to offer most promise as a formative tool that will not only identify students who patients perceive to have weak consultation skills (for their own and faculty attention) but will include guidance on the areas of concern with the addition of powerful free-text comments. Thus, the CARE measure offers both a screening and educational tool, with the advantage of involving patients more in the learning process as advocated by the GMC and others. This will help familiarize students with the growing role of patient feedback on their care, which is rapidly becoming a normal element of healthcare. Finally, the CARE measure fits with the recent changes in the educational programmes of learning outcome, in which communication skills and professionalism are considered to be important approaches in the new learning system.

It should be no surprise that patient feedback is key. This thesis shows that it is also possible, feasible and desirable in undergraduate education. Medical schools should grasp this opportunity. The CARE Measure should be incorporated into medical undergraduate education and medical schools' curricula. It meets students, universities

and most importantly, patients' needs in helping steer students' training and early development of person-centredness in the practice of healthcare. Introducing the CARE measure for senior medical school students will be important to prepare them for future workplace-based assessments in postgraduate medical education. Moreover, using CARE measure during the undergraduate stage for medical students will make them more familiar with training on this measure, which would be useful as it is part of workplace assessments in the Foundation Programme and beyond.

References

- ALBANESE, M. 1999. Rating educational quality: factors in the erosion of professional standards. *Academic Medicine*, 74, 652-658.
- ALBANESE, M. A., SNOW, M. H., SKOCHELAK, S. E., HUGGETT, K. N. & FARRELL, P. M. 2003. Assessing Personal Qualities in Medical School Admissions. *Acad. Med.*, 78, 313-321.
- ANCEL, G. 2006. Developing empathy in nurses: an inservice training program. *Arch Psychiatr Nurs*, 20, 249-57.
- BAILLIE, L. 1996. A phenomenological study of the nature of empathy. *Journal of Advanced Nursing*, 24, 1300-1308.
- BAKER, R. 1990. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *British Journal of General Practice*, 40, 487-490.
- BALINT, M. 1957. *The Doctor, His Patient and The Illness*, London, Tavistock Publications.
- BARRETT-LENNARD, G. 1978. The Relationship Inventory: Later development and adaptations. *JSAS Catalog of Selected Documents in Psychology*, 8, 68.
- BARRETT-LENNARD, G. T. 1962. Dimensions of Therapist Response as Causal Factors in Therapeutic Change. *Psychological Monographs*, 76, Whole Number 562.
- BARRETT-LENNARD, G. T. 1986. The relationship inventory now: Issues and advances in theory, method and use. In: GREENBERG, L. S. & PINSOFF, W. M. (eds.) *The psychotherapeutic process: A research handbook*. New York: Guilford.
- BASU, S., ROBERTS, C., NEWBLE, D. & SNAITH, M. 2004. Competence in the musculoskeletal system: assessing the progression of knowledge through an undergraduate medical course. *Med Educ.*, 38, 1253-60.
- BAYER-FETZER 2001. Essential elements of communication in medical encounters: the Kalamazoo Consensus statement. *Academic Medicine*, 76, 390-393.
- BECKMAN, H., MARKAKIS, K., SUCHMAN, A. & FRANKEL, R. 1994. The Doctor-Patient Relationship and Malpractice: Lessons From Plaintiff Depositions *Arch Intern Med*, 12, 1365-1370.
- BECKMAN, T. J., GHOSH, A. K., COOK, D. A., ERWIN, P. J. & MANDREKAR, J. N. 2004. How reliable are assessments of clinical teaching? A review of the published instruments. *Journal of General Internal Medicine*, 19, 971-977.

- BELLINI, L. M., BAIME, M. & SHEA, J. A. 2002. Variation of Mood and Empathy During Internship. *The Journal of the American Medical Association JAMA*, 287, 3143-3146.
- BELLINI, L. M. & SHEA, J. A. 2005. Mood change and empathy decline persist during three years of internal medicine training. *Academic Medicine*, 80, 164-167.
- BENNETT, D. A. 2001. How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25, 464-469.
- BERG, K., MAJDAN, J. F., BERG, D., VELOSKI, J. & HOJAT, M. 2011a. Medical students' self-reported empathy and simulated patients' assessments of student empathy: an analysis by gender and ethnicity. *Academic Medicine: Journal Of The Association Of American Medical Colleges*, 86, 984-988.
- BERG, K., MAJDAN, J. F., BERG, D., VELOSKI, J. & HOJAT, M. 2011b. A comparison of medical students' self-reported empathy with simulated patients' assessments of the students' empathy. *Medical teacher*, 33, 388-91.
- BERNSTEIN, D. A. & NASH, P. W. (eds.) 2002. *Essentials of Psychology: Annotated Instructor's Edition.*, Boston: Houghton-Mifflin.
- BERTAKIS KD, R. D., PUTNAM SM. 1991. The relationship of physician medical interview style to patient satisfaction. *J Fam Pract*, 32, 175-81.
- BICKLEY, L. S. & SZILAGYI, P. G. 2009. Interviewing and the Health History *Bates' Guide to Physical Examination and History Taking*. 10th ed. Philadelphia: Lippincott Williams & Wilkins
- BIKKER, A. P., MERCER, S. W. & REILLY, D. 2005. A Pilot Prospective Study on the Consultation and Relational Empathy, Patient Enablement, and Health Changes over 12 Months in Patients Going to the Glasgow Homoeopathic Hospital. *Journal Of Alternative And Complementary Medicine (New York, N.Y.)*, 11, 591-600.
- BLAND, J. M. & ALTMAN, D. G. 1997. Statistics Notes: Cronbach's Alpha. *British Medical Journal*, 314, 572.
- BOUD, D. & FALCHIKOV, N. 2006. Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education*, 31, 399-413.
- BOUD, D. & MOLLOY, E. 2012. *Feedback in Higher and Professional Education: Understanding it and doing it well*, London, Routledge.
- BOUD, D. & MOLLOY, E. 2013. Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38, 698-712.

- BOULET, J. R., BEN-DAVID, M. F., ZIV, A., BURDICK, W. P., CURTIS, M., PEITZMAN, S. & GARY, N. E. 1998. High-stakes examinations: What do we know about measurement? Using Standardized Patients to Assess the Interpersonal Skills of Physicians. *Academic Medicine*, 73, S94-96.
- BRENNAN, R. L. 2001. *Generalizability Theory*, New York, Springer-Verlag.
- BRENNAN, R. L. 2011. *Generalizability theory* [Online]. New York: Springer-Verlag. Available: <http://www.education.uiowa.edu/centers/casma/computer-programs> 2011].
- BRITISH MEDICAL ASSOCIATION 2004. *Communication skills education for doctors: an update*, London, British Medical Association.
- BRITISH MEDICAL ASSOCIATION. 2011. *The psychological and social needs of patients: Medical education and training* [Online]. London: British Medical Association. Available: http://www.bma.org.uk/health_promotion_ethics/psychologicalandsocialneeds_ofpatients.jsp?page=6 [Accessed April 2012].
- BROWN, G., MANOGUE, M. & MARTIN, M. 1999a. The validity and reliability of an OSCE in dentistry. *Eur J Dent Educ*, 3, 117-125.
- BROWN, J. B., BOLES, M., MULLOOLY, J. P. & LEVINSON, W. 1999b. Effect of Clinician Communication Skills Training on Patient Satisfaction: A Randomized Controlled Trail. *Ann Intern Med*, 131, 822-829.
- BROWN, T., WILLIAMS, B., BOYLE, M., MOLLOY, A., MCKENNA, L., MOLLOY, L. & LEWIS, B. 2010. Levels of empathy in undergraduate occupational therapy students. *Occupational Therapy International*, 17, 135-141.
- BURKE, S. E. 2008. *The Doctor-Patient Relationship: An Exploration of Trainee Doctors' Views*. Doctor of Philosophy, The University of Birmingham.
- BURNS, D. D. & AUERBACH, A. 1996. Therapeutic empathy in cognitive-behavioural therapy: does it really make a difference? *Frontiers of cognitive therapy*, 135-164.
- BUSCEMI, N., HARTLING, L., VANDERMEER, B., TJOSVOLD, L. & KLASSEN, T. 2006. Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol* 59, 697-703.
- BYLUND, C. L. & MAKOUL, G. 2002. Empathic communication and gender in the physician-patient encounter. *Patient Education And Counseling*, 48, 207-216.
- CENTOR, R. M. 2007. To Be a Great Physician, You Must Understand the Whole Story. *MedGenMed*, 9, 59.

- CENTRE FOR REVIEWS AND DISSEMINATION 2009. Systematic Review: CRD's guidance for undertaking reviews in health care. York, UK.: CRD, University of York.
- CHEN, D., KIRSHENBAUM, D. S., YAN, J., KIRSHENBAUM, E. & ASELTINE, R. H. 2012. Characterizing changes in student empathy throughout medical school. *Medical Teacher*, 34, 305-311.
- CHEN, D. C., LEW, R., HERSHMAN, W. & ORLANDER, J. D. 2007. A Cross-sectional Measurement of Medical Student Empathy *J Gen Intern Med* 22, 1434-1438.
- CHEN, D. C., PAHILAN, M. E. & ORLANDER, J. D. 2010a. Comparing a self-administered measure of empathy with observed behavior among medical students. *Journal of general internal medicine*, 25, 200-2.
- CHEN, D. C. R., PAHILAN, M. E. & ORLANDER, J. D. 2010b. Comparing a self-administered measure of empathy with observed behavior among medical students. *Journal Of General Internal Medicine*, 25, 200-202.
- CHI, M. T. H. 1997. Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *The Journal of Learning Sciences*, 6, 271-315.
- CINAR, N., CEVAHIR, R., SAHIN, S., SÖZERI, C. & KUGUOGLU, S. 2007. Evaluation of the empathic skills of nursing students with respect to the classes they are attending. *Revista Eletrônica de Enfermagem*, 9, 588-595.
- CLYNES, M. P. & RAFFERTY, S. E. 2008. Feedback: An essential element of student learning in clinical practice. *Nurse Education in Practice*, 8, 405-411.
- COHEN, R., REZNICK, R. K., TAYLOR, B. R., PROVAN, J. & ROTHMA, E. A. 1990. Reliability and validity of the objective structured clinical examination in assessing surgical residents *The American Journal of Surgery*, 160, 302-305.
- COLEMAN, K. & MURRAYA, E. 2002. Patients' views and feelings on the community-based teaching of undergraduate medical students: a qualitative study. *Family Practice* 19, 183-188.
- COLLIVER, J. A., CONLEE, M. J., VERHULST, S. J. & DORSEY, J. K. 2010. Reports of the decline of empathy during medical education are greatly exaggerated: a reexamination of the research. *Academic medicine : journal of the Association of American Medical Colleges*, 85, 588-93.
- COLLIVER, J. A., WILLIS, M. S., ROBBS, R. S., COHEN, D. S. & SWARTZET, M. H. 1998. Assessment of empathy in a standardized-patient examination. *Teach Learn Med*, 10, 8-11.
- COOK, D. A. & BECKMAN, T. J. 2006. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *The American Journal of Medicine*, 119, 166.e7-166.e16.

- CORAH NL, O. S. R., BISSELL GD, THINES TJ, MENDOLA P. 1988. The dentist-patient relationship: perceived dentist behaviors that reduce patient anxiety and increase satisfaction. *J Am Dent Assoc*, 116, 73-6.
- COULEHAN, J. L., PLATT, F. W., EGENER, B., FRANKEL, R., LIN, C.-T., LOWN, B. & SALAZAR., W. H. 2001. "Let Me See If I Have This Right ...": Words That Help Build Empathy. *Annals of Internal Medicine*, 135, 221-227.
- COYLE, J. & WILLIAMS, B. 2001. Valuing people as individuals: development of an instrument through a survey of person-centredness in secondary care. *Journal of Advanced Nursing*, 36, 450-459.
- CROCKER, L. & ALGINA, J. 1986. *Introduction to Classical and Modern Test Theory*, Orlando, FL., Harcourt Brace Jovanovich.
- CRONBACH, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- CRONBACH, L. J., GLESER, G. C., NANDA, H. & RAJARATNAM, N. 1972. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*, New York, John Wiley.
- CROSSLEY, J., HUMPHRIS, G. & JOLLY, B. 2002a. Assessing health professionals. *Medical Education*, 36, 800-804.
- DAELMANS, H., VAN DER HEM-STOKROOS HH, HOOGENBOOM RJI, SCHERPBIER AJJA, STEHOUWER CDA & CPM., V. D. V. 2004. Feasibility and reliability of an in-training assessment programme in an undergraduate clerkship. *Med Educ.*, 38, 1270-7.
- DAVIS, C. M. 1990. What is Empathy, and Can Empathy Be Taught? *Physical Therapy* 70, 707-11.
- DAVIS, D. A., MAZMANIAN, P. E., FORDIS, M., HARRISON, R. V., THORPE, K. E. & PERRIER, L. 2006. Accuracy of Physician Self-assessment Compared With Observed Measures of Competence: A Systematic Review. *JAMA*, 296, 1094-1102.
- DAVIS, M. H. 1980. A Multidimensional Approach to Individual Differences in Empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- DAVIS, M. H. 1983a. Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach. *Journal of Personality and Social Psychology*, 44, 113-126.
- DAVIS, M. H. 1983b. Empathic concern and the muscular dystrophy telethon. Empathy as a multidimensional construct. *Pers Soc Psychol Bull*, 9, 223-229.
- DAWOOD, M. 2005. Patient-centred care: lessons from the medical profession. *Emergency Nurse*, 13, 22-28.

- DE HAES, J., OORT, F. J. & HULSMAN, R. L. 2005. Summative assessment of medical students' communication skills and professional attitudes through observation in clinical practice. *Medical Teacher*, 27, 583-589.
- DEVELLIS, R. F. 2006. Classical Test Theory. *Medical Care*, 44 (11 Supp.3), S50-59.
- DI BLASI Z, HARKNESS E, ERNST E, GEORGIU A & J., K. 2001. Influence of context effects on health outcomes: a systematic review. *The Lancet* 357, 757-762.
- DI BLASI, Z., HARKNESS, E., ERNST, E., GEORGIU, A. & KLEIJNEN, J. 2001. Influence of context effects on health outcomes: a systematic review. *The Lancet*, 357, 757-762.
- DISEKER, R. & MICHIELUTTE, R. 1981. An analysis of empathy in medical students before and following clinical experience. *Med Educ.*, 56, 1004-1010.
- DOCHEFF, D. 1990. The feedback sandwich *Journal of Physical Education, Recreation & Dance*, 61, 17-18.
- DOUGLAS, S. P. & NIJSSEN, E. J. 2003. On the use of "borrowed" scales In cross-national research: A cautionary note. *International marketing Review*, 20, 621-642.
- DOWNING, S. & HALADYNA, T. 2004. Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical Education* 38, 327-333.
- DOWNING, S. M. 2003. Validity: on the meaningful interpretation of assessment data. *Medical Education*, 37, 830-837.
- DUNDEE REPORT GMC 2009. Report on Dundee Medical School, University of Dundee. *Quality Assurance of Basic Medical Education*. General Medical Council.
- DURNING, S. J., CATION, L. J., MARKERT, R. J. & N., P. L. 2002. Assessing the Reliability and Validity of the Mini-Clinical Evaluation Exercise for Internal Medicine Residency Training. *Academic Medicine*, 77, 900-904.
- EISENBERG, N. & LENNON, R. 1983. Sex deferences in empathy and related capacities. *Psychological Bulletin*, 94, 100-131.
- EPSTEIN, O., PERKIN, G. D., COOKSON, J., WATT, I. S., RAKHIT, R., ROBINS, A. W. & HORNETT, G. A. W. 2008. Consultation, medical history and record taking. *Clinical Examination With Student Consult Access*. 4th ed. London: Mosby.
- EPSTEIN, R. & STREET, R. J. 2007. Patient-Centered Communication in Cancer Care: Promoting Healing and Reducing Suffering. National Cancer Institute. Publication No. 07-6225. Bethesda, MD.

- EPSTEIN, R. M. 2007. Assessment in Medical Education. *Medical education*, Volume 356, 387-396.
- EPSTEIN, R. M., DANNEFER, E. F., NOFZIGER, A. C., HANSEN, J. T., JOSPE, N., MELDRUM, S. C. & HENSON, L. C. 2004. Comprehensive Assessment of Professional Competence: The Rochester Experiment. *Teaching and Learning in Medicine*, 16, 186-196.
- EPSTEIN, R. M. & HUNDERT, E. M. 2002. Defining and assessing professional competence. *JAMA*, 287, 226-235.
- EVA, K. W., ROSENFELD, J., REITER, H. I. & NORMAN, G. R. 2004. An admissions OSCE: the multiple mini-interview. *Medical Education* 38, 314-326.
- EVANS, B. J., STANLEY, R. O. & BURROWS, G. D. 1993. Measuring medical students' empathy skills. *The British Journal Of Medical Psychology*, 66 (Pt 2), 121-133.
- EVANS, B. J., SWEET, B. & COMAN, G. 1993b. Behavioral assessment of the effectiveness of a communication programme for medical students. *Medical Education*, 27, 344-50.
- FALLOWFIELD, L., JENKINS, V., FAREWELL, V., SAUL, J., DUFFY, A. & EVES, R. 2002. Efficacy of a Cancer Research UK communication skills training model for oncologists: a randomised controlled trial. *Lancet* 359, 650-656.
- FERGUSON, P. 2011. Student perceptions of quality feedback in teacher education. *Assessment and Evaluation in Higher Education*, 36, 51-62.
- FIELDS, S. K., HOJAT, M., GONNELLA, J. S., MANGIONE, S., KANE, G. & MAGEE, M. 2004. Comparisons of nurses and physicians on an operational measure of empathy. *Evaluation & The Health Professions*, 27, 80-94.
- FIELDS, S. K., MAHAN, P., TILLMAN, P., HARRIS, J., MAXWELL, K. & HOJAT, M. 2011. Measuring empathy in healthcare profession students using the Jefferson Scale of Physician Empathy: health provider--student version. *Journal Of Interprofessional Care*, 25, 287-293.
- FISHER, R. A. 1925. *Statistical Methods for Research Workers*, Edinburgh, Oliver and Boyd.
- FJORTOFT, N., VAN WINKLE, L. J. & HOJAT, M. 2011. Measuring empathy in pharmacy students. *American journal of pharmaceutical education*, 75, 109.
- FRAGSTEIN, M. V., SILVERMAN, J., CUSHING, A., QUILLIGAN, S., SALISBURY, H. & WISKIN, C. 2008. UK Consensus Statement on the Content of Communication Curricula in Undergraduate Medical Education. *Medical Education*, 42, 1100-1107.

- FRIEDMAN BEN-DAVID, M. 2000. The role of assessment in expanding professional horizons. *Medical Teacher*, 22, 472-477.
- FROMAN, R. D. & PELOQUIN, S. M. 2001. Rethinking the use of the Hogan Empathy Scale: a critical psychometric analysis. *The American Journal Of Occupational Therapy: Official Publication Of The American Occupational Therapy Association*, 55, 566-572.
- FUNG, C. S. & MERCER, S. W. 2009a. A qualitative study of patients' views on quality of primary care consultations in Hong Kong and comparison with the UK CARE Measure. *BMC Fam Pract*, 10, 10.
- FUNG, C. S. C., HUA, A., TAM, L. & MERCER, S. W. 2009. Reliability and validity of the Chinese version of the CARE Measure in a primary care setting in Hong Kong. *Family Practice*, 26, 398-406.
- FUNG, C. S. C. & MERCER, S. W. 2009b. A qualitative study of patients' views on quality of primary care consultations in Hong Kong and comparison with the UK CARE Measure. *Bmc Family Practice*, 10.
- GMC 2002. *Recommendations on Undergraduate Medical Education*, London, General Medical Council.
- GMC 2003. *Tomorrow's Doctors: Recommendations on Undergraduate Medical Education*, London, General Medical Council.
- GMC 2007. *The Foundation Programme Curriculum*. General Medical Council and Modernising Medical Careers: The Next Steps.
- GMC 2009a. *Tomorrow's doctors: outcomes and standards for undergraduate medical education*, London, General Medical Council.
- GMC 2009b. *Medical Students: professional values and fitness to practice.* , London, General Medical Council.
- GMC 2009c. *Good Medical Practice.*, London, General Medical Council.
- GMC 2010a. *Standards for curricula and assessment systems*, London, General Medical Council.
- GMC 2010b. *Reliability issues in the assessment of small cohorts* London, General Medical Council.
- GMC 2011a. *Assessment in undergraduate medical education: Advice supplementary to Tomorrow's Doctors (2009)*, London, General Medical Council.
- GMC 2011b. *Learning and assessment in the clinical environment: the way forward*, London, General Medical Council.

- GOODCHILD, C. E., SKINNER, T. C. & PARKIN, T. 2005. The value of empathy in dietetic consultations. A pilot study to investigate its effect on satisfaction, autonomy and agreement. *Journal of Human Nutrition & Dietetics*, 18, 181-185.
- GRIFFIN, S. J., KINMONTH, A.-L., VELTMAN, M. W. M., GILLARD, S., GRANT, J. & STEWART, M. 2004. Effect on Health-Related Outcomes of Interventions to Alter the Interaction Between Patients and Practitioners: A Systematic Review of Trials. *ANNALS OF FAMILY MEDICINE*, 2, 595-608.
- GURMAN, A. S. 1977. The patient's perceptions of the therapeutic relationship. In: GURMAN, A. S. & RAZIN, A. M. (eds.) *Effective Psychotherapy: A handbook of research*. New York: Pergamon.
- HALL, J. A., HORGANB, T. G., STEINC, T. S. & ROTERD, D. L. 2002. Liking in the physician-patient relationship. *patient Educ Couns*, 48, 69-77.
- HALPERN, J. 2001. *Detached Concern to Empathy: Humanizing Medical Practice*. , London, Oxford University Press.
- HAMDY, H. 2006. Blueprinting for the assessment of health care professionals. *Clinical Teacher*, 3, 175-179.
- HARDEN, R. M. 2002. Developments in outcome-based education. *Medical Teacher*, 24, 117-120.
- HARDEN, R. M. 2007. Outcome-Based Education: the future is today. *Medical Teacher*, 29, 625-629.
- HARDEN, R. M., CROSBY, J. R., DAVIS, M. H. & FRIEDMAN, M. 1999. AMEE Guide No. 14: Outcome-based education. Part 5 - From competency to metacompetency: a model for the specification of learning outcomes. *Med Teach*, 21, 546-552.
- HARDEN, R. M., STEVENSON, M., DOWNIE, W. & WILSON, G. M. 1975. Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal*, 1, 447-451.
- HARGIE, O., DICKSON, D., BOOHAN, M. & HUGHES, K. 1998. A survey of communication skills training in UK Schools of Medicine: present practices and prospective proposals. *Medical Education*, 32, 25-34.
- HASLAM, N. 2007. Humanising medical practice: the role of empathy. *The Medical Journal of Australia, BMJ Case Notes* 2009, 187, 381-382.
- HATTIE, J. & TIMPERLEY, H. 2007. The power of feedback. . *Review of Educational Research*, 77, 81-112.
- HELMAN, C. 2003. Introduction: the healing bond. In: HELMAN, C. (ed.) *Doctors and patients. An anthology*. Abingdon: Radcliffe Medical Press.

- HELMANN, C. G. 1981. Diseases versus illness in general practice. *J R Coll Gen Pract*, 31, 548-52.
- HEMMERDINGER, J. M., STODDART, S. D. R. & LILFORD, R. J. 2007. A systematic review of tests of empathy in medicine. *BMC Med Educ*, 7, 24.
- HOGAN, R. 1969. Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307-316.
- HOJAT, M. 2007. *Empathy in patient care: antecedents, development, measurement, and outcomes* M Hojat Berlin, Springer.
- HOJAT, M., GONNELLA, J. S., MANGIONE, S., NASCA, T. J. & MAGEE, M. 2003. Physician Empathy In Medical Education And Practice: Experience With The Jefferson Scale of Physician Empathy. *Seminars in Integrative Medicine*, 1, 25-41.
- HOJAT, M., GONNELLA, J. S., MANGIONE, S., NASCA, T. J., VELOSKI, J. J., ERDMANN, J. B., CALLAHAN, C. A. & MAGEE, M. 2002a. Empathy in medical students as related to academic performance, clinical competence and gender. *Medical Education*, 36, 522-527.
- HOJAT, M., GONNELLA, J. S., NASCA, T. J., MANGIONE, S., VERGARE, M. & MAGEE, M. 2002b. Physician empathy: definition, components, measurement, and relationship to gender and specialty. *The American Journal Of Psychiatry*, 159, 1563-1569.
- HOJAT, M., LOUIS, D. Z., MARKHAM, F. W., WENDER, R., RABINOWITZ, C. & GONNELLA, J. S. 2011a. Physicians' empathy and clinical outcomes for diabetic patients. *Academic medicine : journal of the Association of American Medical Colleges*, 86, 359-64.
- HOJAT, M., MANGIONE, S., KANE, G. C. & GONNELLA, J. S. 2005a. Relationships between scores of the Jefferson Scale of Physician Empathy (JSPE) and the Interpersonal Reactivity Index (IRI). *Medical Teacher*, 27, 625-628.
- HOJAT, M., MANGIONE, S., NASCA, T. J., COHEN, M. J. M., GONNELLA, J. S., ERDMANN, J. B., VELOSKI, J. J. & MAGEE, M. 2001. The Jefferson Scale of Physician Empathy: development and preliminary psychometric data. *Educational & Psychological Measurement*, 61, 349-365.
- HOJAT, M., MANGIONE, S., NASCA, T. J., RATTNER, S., ERDMANN, J. B., GONNELLA, J. S. & MAGEE, M. 2004. An empirical study of decline in empathy in medical school. *Medical Education*, 38, 934-941.
- HOJAT, M., SPANDORFER, J., LOUIS, D. Z. & GONNELLA, J. S. 2011c. Empathic and sympathetic orientations toward patient care: conceptualization, measurement, and psychometrics. *Academic Medicine: Journal Of The Association Of American Medical Colleges*, 86, 989-995.

- HOJAT, M., VERGARE MJ, MAXWELL K, BRAINARD G, HERRINE SK, ISENBERG GA, VELOSKI J & JS., G. 2009. The Devil is in the Third Year: A Longitudinal Study of Erosion of Empathy in Medical School. *Academic Medicine* 84, 1182-1191
- HORNSTEN, A., LUNDMAN, B., SELSTAM, E. K. & SANDSTROM, H. 2005. Patient satisfaction with diabetes care. *Journal of Advanced Nursing*, 51, 609-617.
- HORVATH, A. O. & SYMONDS, B. D. 1991. Relation Between Working Alliance and Outcome in Psychotherapy: A Meta-Analysis. *Journal of Counseling Psychology*, 38, No. 2, 139-149, 139-149.
- HOWE, A. 2007. Ally, advocate, authority: strengthening the patient voice in medical education. *The Clinical Teacher*, 4, 170-174.
- HOWIE, J. G., HEANEY, D. J. & MAXWELL, M. 1997. *Measuring Quality in General Practice: Pilot study of a needs, process and outcome measure.* , London, The royal College of General Practitioners.
- HOWIE, J. G., HEANEY, D. J., MAXWELL, M. & WALKER, J. J. 1998. A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. *Family Practice*, 15, 165-171.
- HOWIE, J. G., HEANEY, D. J., MAXWELL, M., WALKER, J. J., FREEMAN, G. K. & RAI, H. 1999. Quality at general practice consultations: cross sectional survey. *British Medical Journal*, 319, 738.
- HSIAO, C. Y., TSAI, Y. F. & KAO, Y. C. 2012. Psychometric properties of a Chinese version of the Jefferson Scale of Empathy-Health Profession Students. *Journal Of Psychiatric And Mental Health Nursing*.
- HULSMAN, R. L., ROS, W. J. G., WINNUST, J. A. M. & BENSING, J. M. 2002. The effectiveness of a computer-assisted instruction programme on communication skills of medical specialists in oncology. *Medical Education*, 36, 125-134.
- HUNSDAHL, J. B. 1967. Concerning Einfühlung (empathy): A concept analysis of its origin and early development. *Journal of the History of the Behavioral Sciences*, 3, 180-191.
- IBM CORP 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY. IBM Corp.
- IRVING, P. & DICKSON, D. 2004. Empathy: towards a conceptual framework for health professionals. *International Journal of Health Care Quality Assurance*, 17, 212-220.

- JACKSON, S. 1992. The listening healer in the history of psychological healing. *Am J Psychiatry*, 149, 1623-32.
- JENKINS, V. & FALLOWFIELD, L. 2002. Can Communication Skills Training Alter Physicians' Beliefs and Behavior in Clinics? *Journal of Clinical Oncology*, 20, 765-769.
- JHA, V., QUINTON, N. D., BEKKER, H. L. & ROBERTS, T. E. 2009. Strategies and interventions for the involvement of real patients in medical education: a systematic review. *Medical Education*, 43, 10-20.
- JOHNSON, A. H. 2001. The Balint Movement in America. *Family Medicine*, 33 174-7.
- KANE, G. C., GOTTO, J. L., MANGIONE, S., WEST, S. & HOJAT, M. 2007. Jefferson Scale of Patient's Perceptions of Physician Empathy: preliminary psychometric data. *Croatian Medical Journal*, 48, 81-86.
- KATAOKA, H. U., KOIDE, N., HOJAT, M. & GONNELLA, J. S. 2012. Measurement and correlates of empathy among female Japanese physicians. *BMC Medical Education*, 12, 48-48.
- KATAOKA, H. U., KOIDE, N., OCHI, K., HOJAT, M. & GONNELLA, J. S. 2009. Measurement of empathy among Japanese medical students: psychometrics and score differences by gender and level of medical education. *Academic medicine* 84, 1192-7.
- KEEN, A. J. A., KLEIN, S. & ALEXANDER, D. A. 2003. Assessing the communication skills of doctors in training: Reliability and sources of error. *Advances in Health Sciences Education*, 8, 5-16.
- KESTENBAUM, R., FARBER, E. A. & SROUFE, L. A. 1989. Individual differences in empathy among preschoolers: Relation to attachment history. *New direction of child development*, 44, 51-64
- KIM, S. S., KAPLOWITZ, S. & JOHNSTON, M. V. 2004. The effects of physician empathy on patient satisfaction and compliance. *Evaluation and Health Profission*, 27, 237-51.
- KIMBERLIN, C. L. & WINTERSTEIN, A. G. 2008. Validity and reliability of measurement instruments used in research *American Journal of Health-System Pharmacy*, 65, 2276-2284.
- KLISZCZ, J., NOWICKA-SAUER, K., TRZECIAK, B., NOWAK, P. & SADOWSKA, A. 2006. Empathy in health care providers – validation study of the Polish version of the Jefferson Scale of Empathy. *Advances in Medical Sciences*, 51, 220-225.
- KLUGER, A. & VAN DIJK, D. 2010. Feedback, the various tasks of the doctor, and the feedforward alternative. *Medical education*, 44, 1166-1174.

- KLUGER, A. N. & DENISI, A. 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- KRUPAT, E., FRANKEL, R., STEIN, T. & IRISH, J. 2006. The four habits coding scheme: Validation of an instrument to assess clinicians' communication behavior. *Patient Education and Counseling*, 62, 38-45.
- KUO, J.-C., CHENG, J.-F., CHEN, Y.-L., LIVNEH, H. & TSAI, T.-Y. 2012. An exploration of empathy and correlates among Taiwanese nurses. *Japan Journal Of Nursing Science: JJNS*, 9, 169-176.
- KURTZ, S. M. 2002. Doctor-patient Communication: Principles and Practices *The Candian Journal of Neurological Sciences*, 29, Supplement 2: S23-S29.
- KURTZ, S. M. & SILVERMAN, J. D. 1996. The Calgary-Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. *Medical Education*, 30, 83-89.
- LA MONICA, E. L. 1981. Construct validity of an empathy instrument. *Research in nursing & health*, 4, 389-400.
- LA MONICA, E. L., CAREW, D. K., WINDER, A. E., HAASE, A. M. & BLANCHARD, K. H. 1976. Empathy training as the major thrust of a staff development program. *Nursing Research*, 25, 447-451.
- LARSON, E. B. & YAO, X. 2005. Clinical Empathy as Emotional Labor in the Patient-Physician Relationship. *The Journal of the Americal Medical Association JAMA*, 293.
- LELORAIN, S., BREDART, A., DOLBEAULT, S. & SULTAN, S. 2012. A systematic review of the associations between empathy measures and patient outcomes in cancer care. *Psychooncology*, 21, 1255-64.
- LEUNG, W. 2002. Competency based medical training: review. *BMJ*, 325, 693-696.
- LEVINSON, W., ROTER, D. L., MULLOOLY, J. P., DULL, V. T. & FRANKEL, R. M. 1997. The Relationship With Malpractice Claims Among Primary Care Physicians and Surgeons. *JAMA*, 277, 553-559.
- LIM, B. T., MORIARTY, H. & HUTHWAITE, M. 2011. "Being-in-role": A teaching innovation to enhance empathic communication skills in medical students. *Medical Teacher*, 33, e663-e669.
- LIPPS, T. 1903. Einfühlung, inner Nachahmung, und Organ-umpfin-dungen [Empathy, inner imitation, and sensations]. *Archiv fur die gesamte Psychologie*, 2, 185-204.

- LIPPS, T. 1905. Das Wissen von fremden Ichen [Knowledge of other egos]. . *Psychologische Untersuchungen*, 4, 694-722.
- LITTLE, P., EVERITT, H., WILLIAMSON, I., WARNER, G., MOORE, M., GOULD, C., FERRIER, K. & PAYNES, S. 2001. Observational study of effect of patient centredness and positive approach to outcomes of general practice consultations. *British Medical Journal*, 323, 908-911.
- LLOYD, M. & BOR, R. 1996. *Communication Skills for Medicine*, New York, Churchill.
- LOWN, B. A., SASSON, J. P. & HINRICHS, P. 2008. Patients as partners in radiology education: An innovative approach to teaching and assessing patient-centred communication. *Academic Radiology*, 15, 425-432.
- LUMSDEN, M. A., BORE, M., MILLAR, K., JACK, R. & POWIS, D. 2005. Assessment of personal qualities in relation to admission to medical school. *Med Educ.* , 39, 240-2.
- MACPHERSON, H., MERCER, S. W., SCULLION, T. & THOMAS, K. J. 2003. Empathy, enablement, and outcome: an exploratory study of acupuncture patients' perceptions. *Journal Of Alternative And Complementary Medicine* 9, 869-876.
- MAGALHÃES, E., SALGUEIRA, A. P., COSTA, P. & COSTA, M. J. 2011. Empathy in senior year and first year medical students: a cross-sectional study. *BMC medical education*, 11, 52.
- MAGUIRE, P., FAIRBAIRN, S. & FLETCHER, C. 1986. Consultation skills of young doctors: I-Benefits of feedback training in interviewing as students persist. *BMJ*, 292, 1573-1576.
- MANGIONE, S., KANE, G. C., CARUSO, J. W., GONNELLA, J. S., NASCA, T. J. & HOJAT, M. 2002. Assessment of empathy in different years of internal medicine training. *Medical Teacher*, 24, 370-373.
- MCCORMACK, B. 2003. A conceptual framework for person-centred practice with older people. *International Journal of Nursing Practice*, 9, 202-209.
- MCGILTON, K., IRWIN-ROBINSON, H., BOSCAR, V. & SPANJEVIC, L. 2006. Communication enhancement: nurse and patient satisfaction outcomes in a complex continuing care facility. . *Journal of Advanced Nursing*, 54, 35-44.
- MCKENNA, L., BOYLE, M., BROWN, T., WILLIAMS, B., MOLLOY, A., LEWIS, B. & MOLLOY, L. 2011. Levels of empathy in undergraduate midwifery students: An Australian cross-sectional study. *Women & Birth*, 24, 80-84.
- MCKENNA, L., BOYLE, M., BROWN, T., WILLIAMS, B., MOLLOY, A., LEWIS, B. & MOLLOY, L. 2012. Levels of empathy in undergraduate nursing students. *International Journal of Nursing Practice*, 18, 246-251.

- MCKIMM, J., BYRNE, A. & DAVIES, H. 2009. Multisource Feedback assessment of Medical Students' professionalism: who should be involved, when and how? *International Journal of Clinical Skills*, 3, 125-133.
- MCMILLAN, L. R. & SHANNON, D. M. 2011. Psychometric Analysis of the JSPE Nursing Student Version R: Comparison of Senior BSN Students and Medical Students Attitudes toward Empathy in Patient Care. *ISRN nursing*, 2011, 726063.
- MCWHINNEY, I. 1985. Patient-centred and doctor-centred models of clinical decision making. In: SHELDON, M., BROOKE, J. & RECTOR, A. (eds.) *Decision making in general practice*. Basingstoke, UK: Macmillan.
- MCWHINNEY, I. 1996. The importance of being different. *British Journal of General Practice*, 46, 433-436.
- MCWHINNEY, I. & FREEMAN, T. 2009. *Textbook of Family Medicine*, 3rd edition. Oxford, Oxford University Press.
- MEAD, N. & BOWER, P. 2000. Patient-centredness: a conceptual framework and review of the empirical literature. *Social Science and Medicine*, 51, 1087-1110.
- MEHRABIAN, A. 1997a. *The Balanced Emotional Empathy Test (BEES) and Optional Software* [Online]. Monterey, CA. Available: <http://www.kaaj.com/psych/scales/emp.html> [Accessed 12 Dec 2010].
- MEHRABIAN, A. 1997b. Relations Among Personality Scales of Aggression, Violence, and Empathy: Validation Evidence Bearing on the Risk of Eruptive Violence Scale. *Aggressive Behavior*, 23, 433-445.
- MEHRABIAN A, YOUNG AL & SATO S. 1988. Emotional empathy and associated individual differences. *Curr Psychol Res Rev*, 8, 221-240.
- MEHRABIAN, A. & EPSTEIN, N. 1972. A measure of emotional empathy. *Journal of personality*, 40, 525-43.
- MERCER, S. W. 2009. *The CARE Measure – summary of research and current use* [Online]. Glasgow. Available: http://www.gla.ac.uk/media/media_141446_en.pdf.
- MERCER, S. W., FITZPATRICK, B., GOURLAY, G., VOJT, G., MCCONNACHIE, A. & WATT, G. C. 2007c. More time for complex consultations in a high deprivation practice is associated with increased patient enablement. *British Journal of General Practice*, 57, 960-966.
- MERCER, S. W., FUNG, C. S., CHAN, F. W., WONG, F. Y., WONG, S. Y. & MURPHY, D. 2011a. The Chinese-version of the CARE Measure reliably differentiates between doctors in primary care: a cross-sectional study in Hong Kong. *BMC Family Practice*, 12, 43.

- MERCER, S. W., FUNG, C. S. C., CHAN, F. W. K., WONG, F. Y. Y., WONG, S. Y. S. & MURPHY, D. 2011b. The Chinese-version of the CARE measure reliably differentiates between doctors in primary care: a cross-sectional study in Hong Kong. *BMC Family Practice*, 12, 43-43.
- MERCER, S. W., HATCH DJ, MURRAY A, MURPHY DJ & KW, E. 2008b. Capturing patients' views on communication with anaesthetists: the CARE measure. *Clinical Governance* 13, 128-37.
- MERCER, S. W. & HOWIE, J. G. 2006. CQI-2 - a new measure of holistic interpersonal care in primary care consultations. *British Journal of General Practice*, 56, 262-268.
- MERCER, S. W., JANI, B. D., MAXWELL, M., WONG, S. Y. & WATT, G. C. 2012. Patient enablement requires physician empathy: a cross-sectional study of general practice consultations in areas of high and low socioeconomic deprivation in Scotland. *BMC family practice*, 13, 6.
- MERCER, S. W., MAXWELL, M., HEANEY, D. & WATT, G. C. 2004. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Family Practice*, 21, 699-705.
- MERCER, S. W., MAXWELL, M., HEANEY, D. & WATT, G. C. M. 2004a. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Family practice*, 21, 699-705.
- MERCER, S. W., MCCONNACHIE, A., MAXWELL, M., HEANEY, D. & WATT, G. C. M. 2005b. Relevance and practical use of the Consultation and Relational Empathy (CARE) Measure in general practice. *Family practice*, 22, 328-334.
- MERCER, S. W. & MURPHY, D. J. 2008. Validity and reliability of the CARE Measure in secondary care. *Clinical Governance: An International Journal*, 13, 269-283.
- MERCER, S. W., NEUMANN, M., WIRTZ, M., FITZPATRICK, B. & VOJT, G. 2008a. Effect of general practitioner empathy on GP stress, patient enablement, and patient-reported outcomes in primary care in an area of high socio-economic deprivation in Scotland: a pilot prospective study using structural equation modelling. *Patient education and counseling*, 73, 240-245.
- MERCER, S. W. & REILLY, D. 2004. A qualitative study of patient's views on the consultation at the Glasgow Homoeopathic Hospital, an NHS integrative complementary and orthodox medical care unit. *Patient Education and Counseling* 53, 13-18.
- MERCER, S. W., REILLY D & WATT GCM 2002b. The importance of empathy in the enablement of patients attending the Glasgow Homoeopathic Hospital. *British Journal of General Practice*.

- MERCER, S. W. & REYNOLDS, W. J. 2002a. Empathy and quality of care. *British Journal of General Practice*, 52, S9-S12.
- MERCER, S. W. & WATT, G. 2007b. The Inverse Care Law: Clinical Primary Care Encounters in Deprived and Affluent Areas of Scotland. *Annals of Family Medicine*, 5 503-510.
- MESSICK, S. 1989. Validity In: LINN RL (ed.) *Educational Measurement*. 3rd ed. New York: American Council on Education and Macmillan.
- MILLER, G. E. 1990. The assessment of clinical skills/competence/performance. *Acad Med*, 65, S63-S67.
- MOLLOY, E. 2010. The feedforward mechanism: A way forward in clinical learning? *Medical Education for General Practice*, 44, 1157-1159.
- MORENO, R. 2004. Decreasing cognitive load for novice students: effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32, 99-113.
- MORRISON, J. 2003. ABC of learning and teaching in medicine, Evaluation. *British Medical Journal*, 326, 385-387.
- MORRISON, J. 2005. How to choose tomorrow's doctors. *Medical Education*, 39, 240-242.
- MURPHY, D. J. 2009a. *General Practice Specialist Training: development of a valid, reliable, feasible and acceptable workplace-based assessment with appropriate educational impact*. M.D, University of Dundee.
- MURPHY, D. J., BRUCE, D. A. & EVA, K. W. 2008. Workplace-based assessment for General practitioners: using stakeholder perception to aid blueprinting of an assessment battery. *Med Educ.*, 42, 96-103.
- MURPHY, D. J., BRUCE, D. A., MERCER, S. W. & EVA, K. W. 2009b. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom *Advances in Health Sciences Education* 14, 219-232.
- MURPHY, D. J., GUTHRIE, B., SULLIVAN, F. M., MERCER, S. W., RUSSELL, A. & BRUCE, D. A. 2012. Insightful practice: a reliable measure for medical revalidation. *BMJ Quality and Safety*, 21.
- NARAYANAN, A., GRECO, M. & CAMPBELL, J. L. 2010. Generalisability in unbalanced, uncrossed and fully nested studies. *Medical Education*, 44, 367-378.
- NEIGHBOUR, R. 1987. *The Inner Consultation*, Lancaster, MTP.

- NEUMANN, M., BENSING, J., MERCER, S. W., ERNSTMANN, N., OMMEN, O. & PFAFF, H. 2009. Analyzing the "nature" and "specific effectiveness" of clinical empathy: a theoretical overview and contribution towards a theory-based research agenda. *Patient Educ Couns*, 74, 339-46.
- NEUMANN, M., EDELHÄUSER, F., TAUSCHEL, D., FISCHER, M. R., WIRTZ, M., WOOPEN, C., HARAMATI, A. & SCHEFFER, C. 2011a. Empathy decline and its reasons: a systematic review of studies with medical students and residents. *Academic medicine* 86, 996-1009.
- NEUMANN, M., SCHEFFER, C., TAUSCHEL, D., LUTZ, G., WIRTZ, M. & EDELHAUSER, F. 2012. Physician empathy: definition, outcome-relevance and its measurement in patient care and medical education. *GMS Z Med Ausbild*, 29, Doc11.
- NEUMANN, M., WIRTZ, M., BOLLSCHEWEILER, E., WARM, M., WOLF, J. & PFAFF, H. 2008. Psychometric evaluation of the German version of the "Consultation and Relational Empathy" (CARE) measure at the example of cancer patients. *Psychotherapie, Psychosomatik, medizinische Psychologie*, 58, 5-15.
- NEUMANN, M., WIRTZ, M., ELFRIEDE, B., MERCER, S. W., MATHIAS, W., JÜRGEN, W. & HOLGER, P. 2007. Determinants and patient-reported long-term outcomes of physician empathy in oncology: A structural equation modelling approach. *Patient Education and Counseling*, 69, 63-75.
- NEWTON, B. W., BARBER, L., CLARDY, J., CLEVELAND, E. & O'SULLIVAN, P. 2008. Is There Hardening of the Heart During Medical School? *Academic Medicine*, 83, 244-249.
- NHSSCOTLAND. 2008. *Better Health, Better Care: Action plane what it means for you* [Online]. Edinburgh: The Scottish Government. Available: <http://www.scotland.gov.uk/Resource/Doc/210642/0055693.pdf>.
- NHSSCOTLAND. 2010. *The Healthcare Quality Strategy for NHSScotland* [Online]. Edinburgh: The Scottish Government. Available: <http://www.scotland.gov.uk/Resource/0039/00398674.pdf>.
- NICOL, D. 2011. Good designs for written feedback. In: SVINICKI, M. & MCKEACHIE, W. (eds.) *Teaching tips: Strategies, research and theories for college and university teachers*. 13th ed. Wadsworth: Cengage Learning.
- NICOL, D. & MACFARLANE-DICK, D. 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.
- NIGHTINGALE, S. D., YARNOLD, P. R. & GREENBERG, M. S. 1991. Sympathy, empathy, and physician resource utilization. *J Gen Intern Med.* , 6, 420-3.

- NORCINI, J. 2000. What should we do about unreliable scores? *Medical Education*, 34, 501-502.
- NORCINI, J. J., BLANK, L. L., DUFFY, D. & FORTNA, G. S. 2003. The Mini-CEX: A Method for Assessing Clinical Skills. *Ann Intern Med*, 138, 476-481.
- NORMAN, G. R., VLEUTEN, C. P. M. V. D. & NEWBLE, D. I. (eds.) 2002. *International handbook of research in medical education*, Dordrecht, The Netherlands: Kluwer Academic Publishers.
- OLINICK, S. 1984. A critique of empathy and sympathy. In: J. LICHTENBERG, B., AND D.SILVER. (ed.) *Empathy*. Hillsdale: NJ: Erlbaum.
- ONG, L. M. L., VISSER, M. R. M., LAMMES, F. B. & DE HAES, J. C. J. M. 2000. Doctor-patient communication and cancer patients' quality of life and satisfaction. *Patient Education and Counseling*, 41, 145-156.
- ORSMOND, P., MAW, S. J., PARK, J. R., GOMEZ, S. & CROOK, A. C. 2013. Moving feedback forward: theory to practice. *Assessment & Evaluation in Higher Education* 38, 240-252.
- OXFORD ENGLISH DICTIONARY 2011. The Oxford advanced dictionary,. Oxford University press.
- OZCAN, C. T., OFLAZ, F. & BAKIR, B. 2012. The effect of a structured empathy course on the students of a medical and a nursing school. *International Nursing Review*, 59, 532-538.
- PARO, H. B. M. S., DAUD-GALLOTTI, R. M., TIBÉRIO, I. C., PINTO, R. M. C. & MARTINS, M. A. 2012. Brazilian version of the Jefferson Scale of Empathy: psychometric properties and factor analysis. *BMC Medical Education*, 12, 73-73.
- PASSAPERUMA, K., HIGGINS, J., POWER, S. & TAYLOR, T. 2008. Do patients' comfort levels and attitudes regarding medical student involvement vary across specialties? *Medical Teacher*, 30, 48-54.
- PEDERSEN, R. 2009. Empirical research on empathy in medicine-A critical review. *Patient education and counseling*, 76, 307-22.
- PELTIER, B. 2001. The psychology of executive coaching: Theory and Application.
- PENDLETON, D., SCHOFIELD T, TATE P & HAVELOCK P. 1984. *The Consultation: an approach to learning and teaching*, Oxford, Oxford University Press.
- PENG, C. Y., HARWELL, M., LIOU, S.-M. & EHMAN, L. H. 2006. Advances in missing data methods and Implications for Educational Research. In: S.SAWILOWSKY (ed.) *Real data analysis*. Greenwich, CT: Information Age.

- PETRUSA, E. R., BLACKWELL, T. A. & AINSWORTH, M. A. 1990. Reliability and Validity of an Objective Structured Clinical Examination for Assessing the Clinical Performance of Residents. *Arch Intern Med*, 150, 373-577.
- PMETB 2007. Developing and maintaining an assessment system - a PMETB guide to good practice. London: Postgraduate Medical Education and Training Board.
- POPAY, J., ROBERTS, H., SOWDEN, A., PETTICREW, M., ARAI, L., RODGERS, M., BRITTEN, N., ROEN, K. & DUFFY, S. 2006. Guidance on the conduct of narrative synthesis in systematic reviews. . *ESRC Research Methods Programme*.
- POUND, P., GOMPERTZ, P. & EBRAHIM, S. 1998. A patient-centred study of the consequences of stroke. *Clinical Rehabilitation*, 12, 338-347.
- PRICE, S., MERCER, S. W. & MACPHERSON, H. 2006. Practitioner empathy, patient enablement and health outcomes: a prospective study of acupuncture patients. *Patient education and counseling*, 63, 239-245.
- QUINCE, T. A., PARKER, R. A., WOOD, D. F. & BENSON, J. A. 2011. Stability of empathy among undergraduate medical students: A longitudinal study at one UK medical school. *BMC Medical Education* 2011, 11:90, 11.
- RAKEL, D., BARRETT, B., ZHANG, Z., HOEFT, T., CHEWNING, B., MARCHAND, L. & SCHEDER, J. 2011. Perception of empathy in the therapeutic encounter: effects on the common cold. *Patient Education And Counseling*, 85, 390-397.
- RAKEL, D. P., HOEFT, T. J., BARRETT, B. P., CHEWNING, B. A., CRAIG, B. M. & NIU, M. 2009. Practitioner empathy and the duration of the common cold. *Family Medicine*, 41, 494-501.
- RAMAPRASAD, A. 1983. On the definition of feedback. *Behavioral Science*, 28, 4-13.
- RAMJAN, L. M. 2004. Nurses and the “therapeutic relationship”: Caring for adolescents with anorexia nervosa. *Journal of Advanced Nursing*, 45, 495-503.
- RCGP. 2011. *iMAP (Membership by Assessment of Performance): the route to RCGP membership for established GPs* [Online]. Royal College of General Practitioners. Available: <http://www.rcgp.org.uk/gp-training-and-exams/membership-by-assessment-of-performance-imap-the-route-to-rcgp-membership-for-established-gps.aspx> [Accessed January 2011].
- RCGP SCOTLAND. 2003. *Revalidation toolkit for doctors working in clinical general practice in Scotland*. RCGP Scotland [Online]. Scotland: Royal College of General Practitioners. Available: <http://www.rcgp-scotland.org.uk> [Accessed October 2010].

- RCGP SCOTLAND. 2006. *The Care Measure* [Online]. Available: www.caremeasure.org/.
- REINDERS, M. E., BLANKENSTEIN, A. H., HORST, H. T. E. V. D., KNOL, D. L., SCHOONHEIM, P. L. & MARWIJK, H. W. J. V. 2010. Does patient feedback improve the consultation skills of general practice trainees? A controlled trial. *Medical Education* 44, 156-164.
- REINDERS, M. E., BLANKENSTEIN, A. H., MARWIJK, H. W. J., SCHLEYPEN, H., SCHOONHEIM, P. L. & STALMAN, W. A. B. 2008. Development and feasibility of a patient feedback programme to improve consultation skills in general practice training. *Patient Education and Counseling*, 72, 12-19.
- RETHANS, J.-J., NORCINI, J., BARON-MALDONADO, M., BLACKMORE, D., JOLLY, B., LADUCA, T., LEW, S., PAGE, G. & SOUTHGATE, L. 2002. The relationship between competence and performance: implications for assessing practice performance. *Medical Education*, 36, 901-909.
- REYNOLDS, W. J. 1994. The influence of clients' perceptions of the helping relationship in the development of an empathy scale. *Journal of Psychiatric and Mental Health Nursing*, 1, 23-30.
- REYNOLDS, W. J. 2000. *The measurement and development of empathy in nursing*, Aldershot, Ashgate Publishing Limited.
- REYNOLDS, W. J. & BRIAN, S. 2000. Do nurses and other professional helpers normally display much empathy? *Journal of Advanced Nursing*, 31, 226-234.
- RIESS, H., KELLEY, J. M., BAILEY, R., KONOWITZ, P. M. & GRAY, S. T. 2011. Improving empathy and relational skills in otolaryngology residents: a pilot study. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery*, 144, 120-2.
- RIESS, H., KELLEY, J. M., BAILEY, R. W., DUNN, E. J. & PHILLIPS, M. 2012. Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. *Journal Of General Internal Medicine*, 27, 1280-1286.
- RINGSTED, C., PALLISGAARD, J., ØSTERGAARD, D. & SCHERPBIER, A. 2004. The effect of in-training assessment on clinical confidence in postgraduate education. *Med Educ.*, 38, 1261-9.
- RODGERS M, SOWDEN A, PETTICREW M, ARAI L, ROBERTS H, BRITTEN N & POPAY J 2009. Testing methodological guidance on the conduct of narrative synthesis in systematic reviews: effectiveness of interventions to promote smoke alarm ownership and function. *Evaluation*, 15, 47-71.
- ROGER, C. (ed.) 1951. *Client-Centered Therapy*, Boston: Mass: Houghton Mifflin Co.

- ROH, M. S., HAHM, B. J., LEE, D. H. & SUH, D. H. 2010. Evaluation of empathy among Korean medical students: a cross-sectional study using the Korean Version of the Jefferson Scale of Physician Empathy. *Teaching and learning in medicine*, 22, 167-71.
- ROSEN, I. M., GIMOTTY, P. A., SHEA, J. A. & BELLINI, L. M. 2006. Evolution of Sleep Quantity, Sleep Deprivation, Mood Disturbances, Empathy, and Burnout among Interns. *Academic Medicine* 81, 82-85.
- ROTER, D., STEWART, M., PUTNAM, S., LIPKIN, M. J., STILES, W. & INUI, T. 1997. Communication patterns of primary care physicians. *JAMA*, 277, 350-356.
- ROTER, D. L., HALL, J. A., KERN, D. E., BARKER, R., COLE, K. A. & ROCA, R. P. 1995. Improving physicians' interviewing skills and reducing patients' emotional distress. A randomized clinical trial. *Arch Intern Med.*, 155, 1877-1884.
- SADLER, R. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18.
- SARGEANT, J., MANN, K. & FERRIER, M. 2005. Making available a mentoring service to support physician feedback, reflection, learning and change, can increase acceptance and use of feedback. *Medical Education*, 39, 497-504.
- SCHAFER, J. L. 1999. Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8, 3-15.
- SCHAFER, J. L. & GRAHAM, J. W. 2002. Missing Data: Our View of the State of the Art. *Psychological Methods*, 7, 147-177.
- SCHIEKIRKA, S., REINHARDT, D., HEIM, S., FABRY, G., PUKROP, T., ANDERS, S. & RAUPACH, T. 2012. Student perceptions of evaluation in undergraduate medical education: A qualitative study from one medical school. *BMC Medical Education*, 12, 45-51.
- SCHLOMER, G. L., BAUMAN, S. & CARD, N. A. 2010. Best Practices for Missing Data Management in Counseling Psychology. *Journal of Counseling Psychology* 57, 1-10.
- SCHOFIELD, P. E., BUTOW, P. N., THOMPSON, J. F., TATTERSALL, M. H. N., BEENEY, L. J. & DUNN, S. M. 2003. Psychological responses of patients receiving a diagnosis of cancer. *Annals of Oncology*, 14, 48-56.
- SCHOUTEN, B. C., EIJKMAN, M. A. & HOOGSTRATEN, J. 2003. Dentists' and patients' communicative behaviour and their satisfaction with the dental encounter. *Community Dent Health.*, 20, 11-15.
- SCHUWIRTH, L. & VAN DER VLEUTEN, C. 2004a. Merging views on assessment. *Med Educ.*, 38, 1208-1211.

- SCHUWIRTH, L. & VAN DER VLEUTEN, C. 2004b. Changing education, changing assessment, changing research? *Med Educ.*, 38, 805–812.
- SCOTTISH EXECUTIVE 2002. *Using care services national care standards : a guide*, Edinburgh.
- SCRIVIN, M. 1967. The methodology of evaluation. . In: R.W. TYLER, GAGNE, R. M. & SCRIVEN, M. (eds.) *Perspectives of Curriculum Evaluation (AERA Monograph Series - Curriculum Evaluation)*. Chicago: Rand McNally.
- SHANAFELT, T. D., WEST, C., ZHAO, X., NOVOTNY, P., KOLARS, J., HABERMANN, T. & SLOAN, J. 2005. Relationship Between Increased Personal Well-Being and Enhanced Empathy Among Internal Medicine Residents. *J Gen Intern Med.*, 20, 559-564.
- SHERMAN, J. J. & CRAMER, A. 2005. Measurement of changes in empathy during dental school. *Journal Of Dental Education*, 69, 338-345.
- SHILLING, V., JENKINS, V. & FALLOWFIELD, L. 2003. Factors Affecting Patient And Clinician Satisfaction With The Clinical Consultation: Can Communication Skills Training For Clinicians Improve Satisfaction? *Psycho-Oncology*, 12, 599-611.
- SHUMWAY, J. M. & HARDEN, R. M. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Medical Teacher*, 25, 569-584.
- SHUTE, V. J. 2008. Focus on formative feedback. *Review of Educational Research*, 78, 153-189.
- SILVERMAN, J. D., KURTZ, S. M. & DRAPER, J. 2005. *Skills for Communicating with Patients (2nd edn)*, Oxford, Radcliffe Publishing.
- SIMPSON, J. G., FURANCE, J., CROSBY, J., CUMMING, A. D., EVANS, P. A., DAVID, M. F. B., HARDEN, R. M., LLOYD, D., MCKENZIE, H., MCLACHLAN, J. C., G.F.MCPHAT, PERCY-ROBB, I. W. & MACPHERSON, S. G. 2002. The Scottish doctor--learning outcomes for the medical undergraduate in Scotland : a foundation for competent and reflective practitioners. . *Medical Teacher*, 24, 136-143.
- SIMPSON, M., BUCKMAN, R., STEWART, M., MAGUIRE, P., LIPKIN, M., NOVACK, D. & TILL, J. 1991. Doctor-patient communication: the Toronto consensus statement. *BMJ* 303, 1385-7.
- SNELL, L., TALLETT, S., HAIST, S., HAYS, R., NORCINI, J., PRINCE, K., ROTHMAN, A. & ROWE, R. 2000. A review of the evaluation of clinical teaching: new perspectives and challenges. *Medical Education*, 34, 862-870.

- SOUTHGATE, L. & GRANT, J. 2004. Principles for an Assessment System for Postgraduate Medical Training. *Postgraduate Medical Education and Training Board*.
- SPENCER, J. 2004. Decline in empathy in medical education: how can we stop the rot? Editorials. *Med Educ.*, 38, 916-920.
- SPIRO, H., CURNEN, M. G. M., PESCHEL, E. & JAMES, D. S. (eds.) 1996. *Empathy and the Practice of Medicine: Beyond Pills and the Scalpel* New Haven: Yale University press.
- SQUIER, R. W. 1990. A model of empathic understanding and adherence to treatment regimens in practitioner-patient relationships. *Social Science & Medicine*, 30, 325-339.
- STEPIEN, K. A. & BAERNSTEIN, A. 2006. Educating for empathy: a review. *J Gen Intern Med*, 21, 524-530.
- STEWART, M. A. 1995. Effective Physician-patient Communication and Health Outcomes: a review. *CAN MED ASSOC J*, 152, 1423-1433.
- STOBER, D. R. & GRANT, A. (eds.) 2006. *Evidence-based coaching handbook: putting best practices to work for your clients.*, Hoboken, NJ: Wiley & Sons.
- STREET, R. L., MAKOUL, G., ARORA, N. K. & EPSTEIN, R. M. 2009. How does communication heal? Pathways linking clinician-patient communication to health outcomes. *Patient Education and Counseling* 74, 295-301.
- STREINER, D. 1985. Global rating scales. In: NEUFELD, V. & NORMAN, G. R. (eds.) *Assessing clinical competence*. New York: Springer.
- STREINER, D. & NORMAN, G. R. 2008. *Health Measurement Scales: A Practical Guide to their Development and Use* Oxford, Oxford University Press.
- SUH, D. H., HONG, J. S., LEE, D. H., GONNELLA, J. S. & HOJAT, M. 2012. The Jefferson Scale of Physician Empathy: A preliminary psychometric study and group comparisons in Korean physicians. *Medical Teacher*, 34, e464-8.
- TAMBURRINO, M. B., LYNCH, D. J., NAGEL, R. & MANGENC, M. 1993. Evaluating empathy in interviewing: Comparing self-report with actual behavior. *Teaching and Learning in Medicine* 5(4), 217-220.
- TARAS, M. 2005. Assessment -Summative and Formative- some theoretical reflections. *British Journal of Educational Studies* 53, 466-478.
- TAVAKOL, S., DENNICK, R. & TAVAKOL, M. 2011a. Psychometric properties and confirmatory factor analysis of the Jefferson Scale of Physician Empathy. *Medical Education*, 11.

- TAVAKOL, S., DENNICK, R. & TAVAKOL, M. 2011b. Empathy in UK medical students: differences by gender, medical year and specialty interest. *Education for primary care : an official publication of the Association of Course Organisers, National Association of GP Tutors, World Organisation of Family Doctors*, 22, 297-303.
- TCZAC, A. W. & SCHWARZ, M. R. 2000. Minimum Essential Requirements and Standards in Medical Education. *Medical Teacher*, 22, 555-559.
- THE SCOTTISH DOCTOR. 2007. *Learning Outcomes for the Medical Undergraduate in Scotland: A Foundation for Competent and Reflective Practitioners* [Online]. The Scottish Doctor. Available: <http://www.scottishdoctor.org/> [Accessed 31.10 2011].
- THOMAS, M. R., DYRBYE, L. N., HUNTINGTON, J. L., LAWSON, K. L., NOVOTNY, P. J., SLOAN, J. A. & SHANAFELT, T. D. 2007. How Do Distress and Well-being Relate to Medical Student Empathy? A Multicenter Study. *J Gen Intern Med* 22, 177-183.
- THORNDIKE, R. L. 1982. *Applied Psychometrics*, Boston, MA., Houghton Mifflin.
- TITCHENER, E. B. 1909. *Lectures on the Experimental psychology of the thought-processes*, New York, Macmillan.
- TOWLE, A. 2006. Where's the patient's voice in health professional education? *Nurse Education in Practice*, 6, 300-302.
- TROCHIM, W. M. K. 2006. *Research Methods Knowledge Base* [Online]. New York Available: <http://www.socialresearchmethods.net/kb/> [Accessed september 2010].
- TUCKETT, D., BOULTON, M., OLSEN, C. & WILLIAMS, A. 1985. *Meetings Between Experts: An Approach to Sharing Ideas in Medical Consultations*, London, Tavistock Publications.
- UNIVERSITY OF DUNDEE. 2010. *Dundee MBChB Course Structure* [Online]. Dundee: University of Dundee. Available: <http://medicine.dundee.ac.uk/dundee-mbchb-course-structure> March, 2010].
- VAN DER VLEUTEN, C. 1996. The assessment of professional competence: Developments, research and practical implications *Advances in Health Sciences Education* 1, 41-67.
- VAN DER VLEUTEN, C. & SCHUWIRTH, L. W. T. 2005. Assessing professional competence: from methods to programmes. *Medical Education*, 39, 309-317.
- VAN DER VLEUTEN, C. P. M., NORMAN, G. R. & DE GRAAF, E. 1991. Pitfalls in the pursuit of objectivity: issues of reliability. *Medical Education*, 25, 110-118.

- VAN DER VLEUTEN, C. P. M., SCHUWIRTH, L. W. T., SCHEELE, F., DRIESSEN, E. W., HODGES, B., CURRIE, R. & CURRIE, E. 2010. The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics and Gynaecology*, 24, 703–719.
- VAN MOOK, W. N., VAN LUIJK, S. J., FEY-SCHOENMAKERS, M. J., TANS, G., RETHANS, J. J., SCHUWIRTH, L. W. & VAN DER VLEUTEN, C. P. M. 2010. Combined formative and summative professional behaviour assessment approach in the bachelor phase of medical school: A Dutch perspective. *Medical Teacher*, 32, e517–e531.
- WALL, D. 2010. Evaluation: improving practice influencing policy. In: SWANICK, T. (ed.) *Understanding medical education: evidence, theory and practice*. West Sussex: Wiley Blackwell.
- WARD, J., CODY, J., SCHAAL, M. & HOJAT, M. 2012. The empathy enigma: An empirical study of decline in empathy among undergraduate nursing students. *Journal of Professional Nursing*, 28, 34-40.
- WASS, V., BOWDEN, R. & JACKSON, N. 2007. *The principles of Assessment Design*. In: *Assessment in Medical Education and Training*, Oxford, Radcliffe publishing.
- WASS, V., VAN DER VLEUTEN, C., SHATZER, J. & JONES, R. 2001. Assessment of Clinical Competence. *THE LANCET* 357, 945-949.
- WASSERMAN, R. C., INUI, T. S., BARRIATUA, R. D., CARTER, W. B. & LIPPINCOTT, P. 1984. Pediatric clinicians' support for parents makes a difference: an outcome-based analysis of clinician-parent interaction. *Pediatrics*, 74, 1047-1053.
- WEST, C. P., HUSCHKA, M. M., NOVOTNY, P. J., SLOAN, J. A., KOLARS, J. C., HABERMANN, T. M. & SHANAFELT, T. D. 2006. Association of perceived medical errors with resident distress and empathy: a prospective longitudinal study. *JAMA*, 296, 1071-1078.
- WHITE, E., ARMSTRONG, B. K. & SARACCI, R. 2009. *Reducing measurement error and its effects* [Online]. Oxford: Oxford University Press. Available: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198509851.001.0001/acprof-9780198509851-chapter-5> 2014].
- WILLIAMS, B., BROWN, T., BOYLE, M. & DOUSEK, S. 2013. Psychometric testing of the Jefferson Scale of Empathy Health Profession Students' version with Australian paramedic students. *Nursing & Health Sciences*, 15, 45-50.
- WIRTZ, M., BOECKER, M., FORKMANN, T. & NEUMANN, M. 2011. Evaluation of the "Consultation and Relational Empathy" (CARE) measure by means of Rasch-analysis at the example of cancer patients. *Patient Educ Couns*, 82, 298-306.

- WISPÉ, L. 1986. The distinction between sympathy and empathy: To call forth a concept, a word is needed. . *Journal of Personality and Social Psychology* 50, 314-321.
- WONG, S. & LEE, A. 2006. Communication Skills and Doctor Patient Relationship. *Medical Bulletin*, 11, 7-9.
- YU, J. & KIRK, M. 2008. Measurement of empathy in nursing research: systematic review. *Journal of advanced nursing*, 64, 440-454.
- YU, J. & KIRK, M. 2009. Evaluation of empathy measurement tools in nursing: systematic review. *Journal of advanced nursing*, 65, 1790-806.
- ZELDOW, P. B. & DAUGHERTY, S. R. 1987. The stability and attitudinal correlates of warmth and caring in medical students. *Med Educ.*, 21, 353-357.

Appendices

Appendix 1: Search procedure for systematic review

Databases used for search:

EBESCO MEDLINE, EBESCO CINAHL, PUBMED, PSYCNET, WEB OF SCIENCE DIRECT (to July 2013)

Search string for identifying articles:

1. Empathy
2. Empathic
3. Empathise
4. 1 or 2 or 3
5. Measure
6. Assessment
7. Scale
8. Rating
9. Evaluation
10. Tool
11. Questionnaire
12. 5 or 6 or 7 or 8 or 9 or 10 or 11
13. 4 AND 12

	MEDLINE	PUBMED	WEB OF SCIENCE	PSYCNET	SCOPUS
Titles reviewed	4300	1909	4226	481	293

Total number of titles identified after duplicate removed = 4228

Abstract reviewed = 615

Papers reviewed = 81

Papers included = 74

Appendix 2: Ethical Approval Letter

Nashwan Alnoman

From: Peter Willatts <p.willatts@dundee.ac.uk>
Sent: 19 October 2009 17:46
To: Nashwan Hanna
Cc: Jon Dowell; Elizabeth Evans
Subject: Ethics application UREC 9060, "A study of the Consultation And Relational Empathy..."
Categories: Red Category

Dear Nashwan,

I have now received comments on your application from our Reviewers, and I have re-read the documents you submitted. I have also taken another look at Bonnie Lynch's earlier application (UREC 7037). My previous view was that your study constituted research, and would therefore need approval by the UREC. However, I am now of the opinion that completion of the CARE questionnaire by patients provides an important method of obtaining feedback on medical student performance during a consultation. This means that the data you collect provide an audit of teaching, and for this reason your work does not require approval from the UREC. If in the future you decide to publish the findings, I will be happy to provide a letter explaining why the study was exempted from ethical review.

One Reviewer commented that it is the parents of any patient under 18 who are going to be asked to fill in the questionnaire, and that this may be inappropriate, as especially teenagers may attend appointments without their parents. We wondered if you might consider lowering the age to something like 16, or perhaps simply excluding patients below the age of 18? If it is somebody else filling in the questionnaire, it may well not be the patient's opinion that is being expressed.

My apologies for the time it has taken to review your application, but it lies on the boundary between research and an audit of teaching, and it has taken some time to decide on which side of the boundary it falls.

Good luck with your work,

Peter Willatts
 Chair, University Research Ethics Committee

Dr Peter Willatts
 School of Psychology, University of Dundee, Nethergate, Dundee, DD1 4HN, UK.
 Email: p.willatts@dundee.ac.uk
 Tel: +44 (0)1382 384618; 384623
 Fax: +44 (0)1382 229993
 The University of Dundee is a registered Scottish charity, No: SC015096

The University of Dundee is a registered Scottish charity, No: SC015096

Appendix 3: The original CARE questionnaire (Doctors' version)

The CARE Measure

© Stewart W Mercer 2004

Please rate the following statements about today's consultation. Please tick one box for each statement and answer every statement.

<i>How was the doctor at ...</i>	Poor	Fair	Good	Very Good	Excellent	Does Not Apply
. Making you feel at ease..... <i>(being friendly and warm towards you, treating you with respect; not cold or abrupt)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Letting you tell your " story" <i>(giving you time to fully describe your illness in your own words; not interrupting or diverting you)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Really listening <i>(paying close attention to what you were saying; not looking at the notes or computer as you were talking)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Being interested in you as a whole person ... <i>(asking/knowing relevant details about your life, your situation; not treating you as "just a number")</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Fully understanding your concerns..... <i>(communicating that he/she had accurately understood your concerns; not overlooking or dismissing anything)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Showing care and compassion.... <i>(seeming genuinely concerned, connecting with you on a human level; not being indifferent or "detached")</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Being Positive..... <i>(having a positive approach and a positive attitude; being honest but not negative about your problems)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Explaining things clearly..... <i>(fully answering your questions, explaining clearly, giving you adequate information; not being vague)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
. Helping you to take control..... <i>(exploring with you what you can do to improve your health yourself; encouraging rather than "lecturing" you)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
0. Making a plan of action with you ... <i>(discussing the options, involving you in decisions as much as you want to be involved; not ignoring your views)</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Appendix 4: The CARE questionnaire (Students' version)

Please be sure you have read the information on the reverse side of this form before answering the questions below.

Tick only one answer for each question. <i>During this consultation, how was the student at...</i>	Poor	Fair	Good	Very Good	Excellent	Does not apply
1. Making you feel at ease? (being friendly and warm towards you, treating you with respect; not cold or abrupt)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
2. Letting you tell “your” story? (giving you time to fully describe your illness in your own words; not interrupting or diverting you)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
3. Really listening? (paying close attention to what you were saying; not looking at the notes or computer as you were talking)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
4. Being interested in you as a whole person? (asking/knowing relevant details about your life, your situation; not treating you as “just a number”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
5. Fully understanding your concerns? (communicating that he/she had accurately understood your concerns; not overlooking or dismissing anything)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
6. Showing care and compassion? (seeming genuinely concerned, connecting with you on a human level; not being indifferent or “detached”)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
7. Being positive? (having a positive approach and a positive attitude; being honest but not negative about your problems)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
8. Explaining things clearly? (fully answering your questions, explaining clearly, giving you adequate information; not being vague)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
9. Helping you to take control? (exploring with you what you can do to improve your health yourself; encouraging rather than “lecturing” you)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
10. Making a plan of action with you? (discussing the options, involving you in decisions as much as you want to be involved; not ignoring your views)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="radio"/>
11. How would you rate the student’s overall behaviour in this consultation?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Do you have any specific comment which might help the student improve his/her consultation skills?

Appendix 5: Instructions to patients

CARE Patient Questionnaire

Please do not write your name on this questionnaire

Thank you very much for agreeing to complete this questionnaire. It should take only about 5 minutes.

The questionnaire asks your opinion about the student's behaviour in the consultation you just had together. We are testing this questionnaire to see whether it could work well as a way of giving feedback to doctors in training.

By completing the questionnaire, you confirm your willingness to have your answers included as part of this research project. Your answers are anonymous and will not be seen by the student or by other staff members, so please feel free to give your honest opinion about this student.

If any question does not apply to the consultation you just had, select "Does not apply" rather than skip the question.

When finished, please seal it in its envelope and return it to a box at the desk or directly to practice manager.

Acknowledgements: The CARE Measure was developed by Dr Stewart Mercer and colleagues as part of a Health Services Research Fellowship funded by the Chief Scientist Office of the Scottish Executive (2000-2003). The intellectual property rights of the measure belong to the Scottish Ministers. The measure is available for use free of charge for staff of the NHS and for research purposes, but cannot be used for commercial purposes. Anyone wishing to use the measure should contact and register with Stewart Mercer (email: stewmercer@blueyonder.co.uk).

© Scottish Executive 2000-2003

Appendix 6: Feedback results to students (email) 2nd & 3rd years

Hi (student name),

Thank you for your participation in CARE study. Please find below your mean score with details of each question. You could compare your result with the average means and confidence intervals (Attached file) which was obtained from the former students who had previously participated in the study.

Code	ST. NAME	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	TOTAL AVG
**	*** **	4.86	4.71	4.68	4.71	4.64	4.71	4.71	4.61	4.68	4.75	4.86	4.72

I invited you to read the following patients' comments.

1. SHE WAS GENUINELY LOVELY AND VERY HELPFUL.
2. JUST ONE. CHECK THAT THE PATIENT IS NOT SLIGHTLY DEAF, IF SO, SPEAK MORE SLOWLY.
3. KEEP UP THE GOOD WORK.
4. SHE'S VERY FRIENDLY, GREAT LISTENER AND WILL MAKE AN EXCELLENT GP.
5. NO, SHE WILL MAKE A BRILLIANT DOCTOR, VERY NICE.

If you have any enquiries please do not hesitate to contact on below email addresses.

Best wishes,

Principle investigator

Appendix 7: Feedback results (Means and Confidence Intervals)

CARE Questionnaire Average Scores and 95% Confidence Intervals for 38 students			
Questions	Means		
1. Making you Feel at ease?	< 4.278	4.61	>4.944
2. Letting you tell your story?	< 4.223	4.59	>4.966
3. Really listening?	< 4.217	4.58	>4.954
4. Being interested in you as whole?	< 4.046	4.52	>4.996
5. Understanding your concerns?	< 4.095	4.51	>4.936
6. Showing care and compassion?	< 4.187	4.57	>4.962
7. Being positive?	< 4.084	4.53	>4.979
8. Explaining things clearly?	< 4.009	4.47	>4.927
9. Helping you to take control?	< 3.881	4.39	>4.9
10. Making plan of action with you?	< 3.864	4.42	>4.976
11. Overall behaviour?	< 4.309	4.69	>5.00
TOTAL AVERAGES?	< 4.144	4.54	>4.928

Note: The values of the above table were calculated from means of 38 students only i.e. participants of academic year 2009-2010 and 2010-2011.

Appendix 8: Dundee GP block min-Clinical Evaluation Exercise (mini-CEX)

Dundee GP Attachment Mini-Clinical Evaluation Exercise (Mini-CEX)

Student's Name _____ GP Tutor _____

Clinical Setting:	In-patient	Acute Admission	GP surgery	Home visit
New or F/up:	New	Follow up	Assessor Grade: GP Other (Please state)	
Focus of Clinical Encounter:				
History	Examination	Diagnosis	Management	Explanation
No. of times patient seen before by student:			0	1-4 5-9 >10
No. of previous mini-CEX by assessor with <u>any</u> student:			0	1-4 5-9 >10

Please grade the following skills using the scale below	Standard for GP block completion						
	Below Standard		Borderline	Meets Standard	Above Standard		U/C
	F	E	D	C	B	A	
History Taking							
Physical Examination							
Communication							
Clinical Judgement							
Professionalism							
Organisation/ Efficiency							
Overall Clinical Care							

* See descriptors of a satisfactory student

** Unable to comment

Examples of good practice	Suggestions for Development
Agreed action:	

	Low									High
Trainee satisfaction with mini-CEX	1	2	3	4	5	6	7	8	9	10
Assessor satisfaction with mini-CEX	1	2	3	4	5	6	7	8	9	10
Time taken in observation (mins):		Time taken for feedback (mins):								
Assessors Signature						Date.....				
Assessors Name										

Appendix 9: List of Positive Patients' Comments

Category	Positive comments
Specific comments	<ul style="list-style-type: none"> • <i>'Very impressed with active listening skills and said when did she didn't know something'</i> • <i>'This student was very good put me at ease listened to me'</i> • <i>'Student made me feel really comfortable'</i> • <i>'Was very understanding and listened closely to my concerns'</i> • <i>'Very nice and understanding took everything into account and checked all'</i> • <i>'Very understanding and listened carefully'</i> • <i>'Very pleasant manner and very understanding'</i> • <i>'She's very friendly, great listener and will make an excellent GP'</i> • <i>'She was very understanding and very helpful'</i> • <i>'Very lovely lady- good at listening and made you feel welcome'</i> • <i>'Good listener'</i> • <i>'Lovely attitude. Relayed back my concerns by actively listening'</i> • <i>'Good understanding of my medical problems and describe my concerns well to supervising GP'</i> • <i>'I was very satisfied that she gave me the best attention / treatment'</i> • <i>'She was very kind caring and listened to everything she will be a first class doctor'</i> • <i>'Very good great listener and sorted everything out good luck in future'</i> • <i>'It all went smoothly. We were seen very promptly and she was very friendly and listened to us (self/child) carefully'</i> • <i>'Felt really listened too. Very approachable and pleasant. Thank you'</i> • <i>'He paid attention and listened carefully. Hope that this will not change when he will graduate after graduation i.e. Trained doctors sometimes do not seem to show care - compassion as much'</i> • <i>'This student was excellent, and builds up a good rapport, taking time to listen'</i> • <i>'Very understanding. Made me feel at ease'</i> • <i>'Good manner and happy to listen, thanks'</i> • <i>'Very pleasant young student who shows a good listening skill'</i>

	<ul style="list-style-type: none"> • <i>'He listened to my concerns and gave me appropriate advice very kindly'</i> • <i>'He really does his best to look after his patients and make them feel comfortable and satisfied. Best of luck to him in the future. He will make a great doctor'</i> • <i>'Good at listening to specific complications from major hip surgery operation. Very helpful'</i> • <i>'Very good listener'</i> • <i>'Generally very good, was sympathetic to my problem and suggest positive direction'</i> • <i>'Was excellent, her manner and compassion was so good she had me in tears! That's what i needed to open up and let my feelings out. She will be a first class doctor'</i> • <i>'Very nice manner which put me at ease. Very thorough and helpful'</i> • <i>'Made me feel at ease + not stupid as i felt that before i went in'</i> • <i>'Student was very good at putting me at ease and asking relevant questions'</i> • <i>'Very professional approach makes me feel at ease'</i> • <i>'The student was very easy to talk to and reached to things in a positive manner making me feel comfortable to talk about even embarrassing things'</i> • <i>'I would suggest that she get as much experience as possible and sure she will make an excellent doctor'</i> • <i>'Very friendly and warm manner. Made me feel at ease and feel confident'</i> • <i>'Extremely nice and profession, made me feel completely at ease'</i> • <i>'Katrina is a credit to you. Great attitude, great personality put me at ease well done. Thanks'</i> • <i>'She was excellent and kind and considerate as i was feeling very emotional. A doctor in the making'</i> • <i>'Has the ability to make you feel at ease she has a polite safe manner. An honest approach. Was excellent. Wish is will with her final studies and best wishes in any future appointment, a caring doctor. Well done'</i> • <i>'Really happy with the way. Claire made me feel and understood and helped me today. Thank you'</i> • <i>'I felt at ease and very comfortable during the consultation'</i> • <i>'The student seemed to empathetic'</i>
--	---

	<ul style="list-style-type: none"> • <i>'Student doctor was friendly and put me at ease. She should keep this attitude as i felt the consultation was informal. The student will make an excellent doctor'</i> • <i>'She put me at ease which made it a satisfactory appointment'</i> • <i>'He was excellent. Very attentive and felt at ease. Top marks'</i> • <i>'Made me feel comfortable and appeared interested in me as a person. Did not rush me'</i> • <i>'He was very helpful and pleasant. Put us at ease of friendliness'</i> • <i>'Very friendly, made me feel at ease'</i> • <i>'He was very well and explained everything well. Hope he does well as a doctor'</i> • <i>'He was excellent, kind and caring. Hope he keeps it up over the years'</i> • <i>'Very thorough and good bedside manner'</i> • <i>'I couldn't fault her she was very attentive and i felt in good hands. Thank you'</i> • <i>'Very friendly and caring. Seemed genuinely interested'</i> • <i>'The doctor was incredibly pleasant and made my child very much at ease'</i> • <i>'She was very thorough in her examination'</i> • <i>'Very good dealt with both me and my son very well'</i> • <i>'Calm and reassuring'</i> • <i>'This young man showed a genuine interest during the consultation, and had a warm and friendly manner. I wish him luck'</i> • <i>'I found the student to be very approachable and knowledgeable. A good bedside manner'</i> • <i>'Very competent-felt well cured for'</i> • <i>'Knew what she was relating to on the illness very good'</i> • <i>'I think you will make a great doctor. Put me at ease during an embarrassing consultation'</i> • <i>'I am a retired GP and highly impressed with her degree of competence. She will make an excellent doctor. Think she will make a very good doctor'</i> • <i>'Very approachable would make a great doctor. Open, warm and friendly, good attributes'</i> • <i>'A very friendly and easy to talk too. Will make an excellent doctor'</i> • <i>'I can see no reason for this young lady to improve. I</i>
--	--

	<p><i>thought she was excellent and a very nice person. She will make a very good doctor`</i></p> <ul style="list-style-type: none"> • <i>'I cannot think of any way the student can improve. I was impressed by her attitude and quite surprised when she summarised and reflected back to me what i had been telling her. This doesn't usually happen`</i> • <i>'I found it helpful because it clarified to me what I'd said and caused me realise something i had not mentioned, or may have been misunderstood`</i> • <i>'Very good with children. Trust her advice`</i> • <i>'The student's interpersonal skills are excellent`</i> • <i>'Has great potential`</i> • <i>'Learn as much as possible from Dr Glinfield. He is a fantastic doctor! The student was very welcoming, maintain as confidence grows. Thanks`</i> • <i>'Excellent communication and understanding`</i> • <i>'Very approachable and easy to communicate with`</i> • <i>'Has a very caring and understanding attitude .10/10`</i> • <i>'Very helpful and easy to communicate with`</i> • <i>'She met and introduces herself to me. Was warm and friendly. Excellent inter-personal skills. Was assured after consultation`</i> • <i>'A very pleasant interview. Easy to talk to and asked good relevant questions`</i> • <i>'Communication skills excellent`</i> • <i>'Made me feel at ease so enabling me to discuss all worries and aspects of case`</i> • <i>'Spoke well and actually took the time to listen to everything. Had to say 10.10 for communication skills and making me feel at ease (thank you)`</i> • <i>'First class communicator with a splendid bright and positive attitude`</i> • <i>'He has a very polite manner and has good communication skills`</i>
Non-specific comments	<ul style="list-style-type: none"> • <i>'He was lovely`</i> • <i>'Super lad`</i> • <i>'Find him very good`</i> • <i>'Thought she did very` well</i> • <i>'She was really good`</i> • <i>'Very good`</i> • <i>'Thought he was very good`</i>

	<ul style="list-style-type: none"> • <i>'She is already doing an excellent job. Thank you'</i> • <i>'Very professional individual'</i> • <i>'No improvement required. Student doctor dealt with things very well'</i> • <i>'She is excellent'</i> • <i>'Should do well'</i> • <i>'He did very well'</i> • <i>'Very good'</i> • <i>'No advice necessary as she was very good all round'</i> • <i>'No special comment. First class student 100 per cent'</i> • <i>'Really good'</i> • <i>'Excellent manner'</i> • <i>'I had no concerns, i thought she was excellent'</i> • <i>'Consultation was superb'</i> • <i>'Excellent service many thanks'</i> • <i>'Completed her consultation very well'</i> • <i>'Found him very reassuring and helpful easy to speak to'</i> • <i>'Very good, helpful manner and approach'</i> • <i>'Found him to be a very polite well informed young man. Very good with babies'</i> • <i>'I would have no problem with the student being my doctor i found her to be very helpful and interested in me as a patient i am sure she will do very well'</i> • <i>'She was very good indeed'</i> • <i>'Excellent service'</i> • <i>'She was superb'</i> • <i>'She was very excellent'</i> • <i>'Very good manner'</i> • <i>'First class'</i> • <i>'Good session'</i> • <i>'He was very good'</i> • <i>'I found this student's skills very good'</i> • <i>'Very good'</i> • <i>'Extremely pleasant person'</i> • <i>'Was very helpful'</i> • <i>'He was very helpful and polite'</i> • <i>'Very friendly and helpful'</i> • <i>'A lovely and pleasant young lady'</i> • <i>'She was helpful and understanding'</i> • <i>'Excellent manner. Shows understanding'</i> • <i>'Very understanding, nice manner'</i>
--	--

	<ul style="list-style-type: none"> • <i>'I felt extremely comfortable in her company'</i> • <i>'Felt very comfortable, good approach'</i> • <i>'I felt the doctor had a genuine concern for my wellbeing'</i> • <i>'Felt very comfortable'</i> • <i>'Great rapport'</i> • <i>'Her manner was excellent with my 2 years old daughter'</i> • <i>'No problems, very attentive, nice manner'</i> • <i>'Very kind no constructive comment required'</i> • <i>'Lovely warm manner'</i> • <i>'Brilliant consultation, felt comfortable'</i> • <i>'Very good, kind and compassionate student'</i> • <i>'Very nice manner'</i> • <i>'Not only very good but not afraid to ask for a doctor's advice'</i> • <i>'Very friendly, very easy to talk'</i> • <i>'Very polite young lady'</i> • <i>'She was very relaxed and friendly'</i> • <i>'Nice, friendly and attentive'</i> • <i>'Very pleasant, helpful and would see her again'</i> • <i>'Very friendly and helpful'</i> • <i>'She was very friendly and pleasant. I wouldn't mind her as my doctor'</i> • <i>'Very good, friendly approach'</i> • <i>'Lovely friendly person'</i> • <i>'She was extremely helpful'</i> • <i>'Very pleasant'</i> • <i>'He was very good and help me'</i> • <i>'She was genuinely lovely and very helpful'</i> • <i>'Very pleasant, calm approach'</i> • <i>'My son said she was very nice and polite'</i> • <i>'Student was very good, friendly'</i> • <i>'Very helpful and positive'</i> • <i>'Very friendly not condescending'</i> • <i>'Excellent and very patient friendly'</i> • <i>'Think she was lovely'</i> • <i>'First class. Lovely girl'</i> • <i>'Very lovely woman'</i> • <i>'The student doctor was very friendly and helpful during the consultation'</i> • <i>'Very pleasant doctor'</i> • <i>'Very polite and friendly'</i>
--	---

	<ul style="list-style-type: none"> • <i>'Very approachable and pleasant. I enjoyed meeting him'</i> • <i>'He is very polite and enjoyed talking to him'</i> • <i>'Really lovely'</i> • <i>'She was very lovely. Nicer than my doctor at home'</i> • <i>'Very helpful and reassuring'</i> • <i>'Very helpful, very pleased'</i> • <i>'Very friendly towards both my kids and myself'</i> • <i>'Very helpful, lovely person, very good with children'</i> • <i>'A very bright and helpful student. A pleasure to be attended by her'</i> • <i>'Very lovely young man. Generally very happy'</i> • <i>'No further comment apart from i was very impressed and she will be a super GP! I wish her all the best'</i> • <i>'The interview went very well. She will make a grand doctor and i wish her the best of luck'</i> • <i>'She will be a fin doctor'</i> • <i>'Excellent will make a great doctor'</i> • <i>'He will make a very good doctor'</i> • <i>'Will make an excellent doctor'</i> • <i>'I think he will make a good doctor'</i> • <i>'Practice will make a very positive doctor'</i> • <i>'No, she will make a brilliant doctor, very nice'</i> • <i>'Will make a great doctor'</i> • <i>'In my opinion the student was an excellent chap and he will make an excellent doctor and be a credit to his profession. I say this with my experience as a social worker'</i> • <i>'She will make an excellent doctor. She has it all now'</i> • <i>'Very good student. Make a good doctor'</i> • <i>'This student has the makings of being a very good doctor'</i> • <i>'Was very happy with my consultation'</i> • <i>'Very impressed thanks'</i> • <i>'I found this student to be 100% committed to the consultation and i feel very relaxed throughout'</i> • <i>'I was very impressed'</i> • <i>'Very relax consultation, excellent'</i> • <i>'I was quite happy with the consultation'</i> • <i>'Best doctor I've seen in ages'</i> • <i>'I couldn't find any fault with her'</i> • <i>'Consultation was great. Lots of concern shown. Very pleasant manner'</i>
--	---

	<ul style="list-style-type: none"> • <i>'I would have no problem seeing this doctor again'</i> • <i>'Consultation was carried out very well'</i> • <i>'Had i not been told the Dr was a student I'd never have known, i felt very confident and comfortable'</i> • <i>'Was very pleased with his approach'</i> • <i>'No apparent faults on areas for improvement noted'</i> • <i>'Was very impressed by all aspects of the interview with him'</i> • <i>'Completely satisfied'</i> • <i>'Carry on as they are, very nice and friendly'</i> • <i>'Continue with confident supportive manner'</i> • <i>'Sincerely hope this lady goes far wish her all success'</i> • <i>'Good luck in your profession'</i> • <i>'Continue as you are and you will make an excellent doctor'</i> • <i>'Keep good work. Was very happy with this manner'</i> • <i>'Just carry on as he did with me'</i> • <i>'Keep up the good work'</i> • <i>'Lovely girl. She should go for'</i> • <i>'Continue as you are'</i> • <i>'Keep up the good work'</i> • <i>'Carry on and he will attain his ambition'</i> • <i>'Just keep doing what you are doing (:'</i> • <i>'Confident and very helpful'</i> • <i>'Very pleasant and confident. Wouldn't hesitate to see him again'</i>
--	---

Appendix 10: List of Negative Patients' Comments

Specific category	Negative comments
Confidence	<ul style="list-style-type: none"> • <i>“Possibly try to sound more confident in her diagnosis. But generally excellent”</i> • <i>“More self-confidence”</i> • <i>“Have a little more confidence in what you say”</i> • <i>“Helps if she's more confident and convincing”</i> • <i>“Kind, caring but needs to sound more confident - to give the patient re-assurance of her knowledge. This will come with experience”</i> • <i>“Display more confidence in assessing the condition”</i> • <i>“Be more positive and believe what you're saying and yourself, but overall very good”</i> • <i>“Could be more confident in her ability. This would give the patient reassurance”</i> • <i>“To be more confident. Lovely person”</i> • <i>“A little quite/ lacking confidence”</i> • <i>“Only criticism is that she was very quietly spoken and lacked confidence a little, which is understandable and easily fixed”</i> • <i>“Being more confident and assertive”</i> • <i>“A little timid, as one might expect, but with an air of competence that inspired confidence. Generally very good”</i> • <i>“Being more confident”</i>
Communication	<ul style="list-style-type: none"> • <i>“In dealing with the elderly with hearing impairment remembering to speak slowly and clearly”</i> • <i>“Could ask the slowdown, when explaining symptoms so could take note accurately”</i> • <i>“Just one. Check that the patient is not slightly deaf, if so, speak more slowly”</i> • <i>“Speaking more clearly, not as bad thing but too soft spoken for those who do not hear well”</i> • <i>“Very pleasant and friendly. Asked twice if i had a cough which suggested she hadn't paid attention to my first answer, but very good apart from that!”</i> • <i>“Please listen carefully to signs, symptoms and concerns. Took a very egalitarian approach to the consultation”</i>

History and examination	<ul style="list-style-type: none"> • <i>“Limited at first but not having the right equipment, but was good once he could see in my eye properly”</i> • <i>“Should heat hands first”</i> • <i>“May be explaining why she was feeling my stomach, kidneys, etc. Try not to lead patient into saying something which isn't the problem (twice had to say that i didn't have acid reflux problems)”</i> • <i>“I did feel she said sorry too frequently, while examining me even i thought i reassured her she wasn't hurting me. Otherwise she was excellent”</i> • <i>“She did use the wrong name (surname instead of first) for my son during consultation but only little mistakes, otherwise she was lovely”</i> • <i>“May be asking more questions about the symptoms experienced. Further explain the cause of the problem and how to prevent it in the future”</i> • <i>“Double check past history of reoccurrence of problem”</i>
Professionalism	<ul style="list-style-type: none"> • <i>“Could be a little more assertive”</i> • <i>“I found the student doctor good but need shadow Dr Abercrombie to add more experience”</i> • <i>“She was maybe a little "overfriendly". I didn't have a problem with this and liked her enormously but some people may have (wrongly) assumed this indicated a lack of professionalism. On the other hand, this could have been because i wasn't ill and felt my consultation was more of a formality than a consultation with other patients might be”</i> • <i>“Little understanding and no willing to understand”</i> • <i>“Very good with a small child which can be very difficult. Was she right approach to people by being empathetic but without going over the top?”</i> • <i>“The student was great but a 10 min appointment became an hour experience at the surgery”</i> • <i>“I got the experience that if he was unsure of something then he should ask a senior member of staff”</i> • <i>“Dress code could be smarted and more appropriate for the surgery, otherwise very satisfied”</i> • <i>“Stop hitching up skirt during consultation, distracting”</i> • <i>“Not to yawn three times during a consultation”</i>

Nervousness	<ul style="list-style-type: none"> • <i>“Just be less nervous as she has no need, she was really good and helpful”</i> • <i>“She seemed slightly nervous, but I’m sure she’ll get over that”</i> • <i>“Don’t be scared, most patients don’t bite”</i> • <i>“Very kind gentle soul could perhaps be a little stronger”</i> • <i>“This student seemed a little nervous, but overall very good and in June she will be a credit to her profession, a very satisfied patient”</i> • <i>“Relax”</i>

Appendix 11: Email of CARE mapping exercise to participants**Phase 3 assessment mapping exercise**

Dear colleague,

Email on behalf of Teaching Dean

We would be grateful if you could spare 10-15 minutes on this Faculty supported online Phase 3 assessment mapping exercise. This is to establish how well we currently assess the GMC 2009 “Tomorrow’s Doctors” outcomes across the clinical years and specifically to inform the application of a possible patient based measure (the CARE questionnaire). Your opinions would be much appreciated and will contribute to the design of assessments within the new Preparation for Practice.

Direct Link to it: <https://www.tipportfolio.co.uk/private/tipp/tipplogin.aspx>

Username: *****

Password: *****

Many thanks,

Jon Dowell (supervisor)
Nashwan Alnoman (PhD student)

Gary Mires
Teaching Dean


Appendix 12: The CARE mapping exercise

[illegible]

[illegible]

Appendix 13: Instructions on cover page of mapping exercise

Dundee Phase 3 - Assessment Tools Mapping Exercise

The general purpose of the exercise is to identify participants' perceptions and experience of the capacity of tools to give feedback on the different GMC Tomorrow's Doctors Outcomes. This will allow us to identify which tools are thought to best test the different outcomes. It will also allow us to identify outcomes for which there is currently inadequate testing. If you are unsure of any of these tools just click on the  to see a description. We want to understand your perception, prior to using all/some of these tools, so even if you have not had experience of them all please still give us your perceptions of each.

Instructions: Enter in each drop down menu a score of 1-7 (1 poor; 3 fair; 5 good; 7 excellent;) of your perception of how well each Dundee Learning Outcome is likely to be tested by each tool. Once you have pressed submit your views will be captured and you will not be able to make any changes.

We strongly suggest that you complete this by columns rather than by rows, simpler and quicker.

Hover Button Text:

Year 4 EMI: Extended Matching Item format test designed to assess the clinical application of knowledge base.

Year 4 OSCE: 12 station clinical OSCE including consultations, examinations, procedures and data interpretation.

5th Year Case Report: Seven outcome focused (E.g. Ethics or Patient Management), Case based essays marked for portfolio.

5th Year Block Feedback Form: Scottish Doctor 12 outcome based form completed by head tutors at the end of each clinical attachment.

Final Year OSCE: Seven 15 minute station clinical OSCE orientated towards role of FY1 – Procedures, consulting, prescribing etc.

Portfolio Exam: Final year assessment of collated written portfolio, grades for all assessed work; clinical attachments etc and 40 minute viva.

Progress Test: Online knowledge based test requiring free text responses to around 250 items.

CARE: An 11 item patient completed form collected following 25 student led consultations in general practice.

Mini CEX: A structured observation and feedback form used to guide evaluation of student's consultation and / or examination skills.

Case Based Discussion: A structured feedback form used to guide assessment of discussion regarding a patient seen in general practice.

DOPS: A structured observation and feedback form used to guide evaluation of an observed procedure such as IV cannulation.

